

XX Venter JC, Adams M, Li PWD, Myers EW;  
XX WP1: 2001-656860/75.  
DR P-PSDB: ABB71957.

PT New isolated nucleic acid detection reagent for detecting 1000 or more  
XX genes from Drosophila and for elucidating cell signalling and cell-cell  
XX interactions -  
PT

PS Claim 1: SEQ ID NO 42662; 21bp + Sequence Listing; English.  
XX

CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL101840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pat\_sequences.

SQ Sequence 4757 BP; 1093 A; 1236 C; 1150 G; 1278 T; 0 other;

Query Match 0.7%; Score 20; DB 23; Length 4757;  
Best Local Similarity 100.0%; Pred. No. 57;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1869 ACAGCGACGAAAAAAGCAG 1888  
Db 2193 ACAGCGACGAAAAAAGCAG 2174  
|||||

RESULT 45  
ID AAX23519/C  
XX AAX23519 standard; DNA: 44453 BP.  
XX  
XX AAX23519;  
DT 23-JUN-1999 (first entry)  
XX  
XX Human kidney aminopeptidase P genomic DNA fragment 3.  
CE  
XX Aminopeptidase; human; Amp; gene therapy; treatment; Amp-deficiency;  
KW prenatal diagnosis; angioedema; antihypertensive agent; atherosclerosis;  
RW arterial stenosis; industrial protein feed; malabsorption syndrome;  
XX proteaceous waste degradation; additive; immunohistochemistry; ss.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO9911799-A2.  
PN  
XX 11-MAR-1999.  
DD  
XX 02-SEP-1998; 98MO-US18426.  
PE  
XX 02-SEP-1997; 97US-0057854.  
PR  
XX (MED1-) MEDICAL COLLEGE GEORGIA RES INST.  
PA  
XX Ryan JW, Sprinkle TTC, Venema RC.  
P- WP1: 1999-205193/17.  
DR  
XX Nucleic acid encoding human aminopeptidase P  
XX  
XX Claim 13: Page 139-165; 201pp; English.  
XX  
XX This invention describes the isolation of a novel human aminopeptidase P  
CC (Amp). This protein is used to produce recombinant Amp and can be used  
CC for gene therapy for treating Amp-deficiency conditions. Its fragments  
CC are used as primers and probes to identify patients with homozygous and

CC heterozygous Amp deficiency, including prenatal diagnosis (patients  
CC defective in Amp are at risk of developing angiodema if treated with  
CC angiotensin-converting enzyme inhibitors), also as antisense inhibitors  
CC in cases of excessive Amp expression. The product of the invention is  
CC also used to identify Amp-expressing sequences in other animals and to  
CC generate transgenic animals, and comparisons of genomic sequences are  
CC used to detect mutations. Amp inhibitors are potentially useful as  
CC antihypertensive agents and to prevent or treat arterial (re)stenosis  
CC or atherosclerosis. The structure of Amp is used to design synthetic  
CC substrates, e.g. for use in Amp assays. Amp, which hydrolyses N-terminal  
CC amino bonds, can be used to degrade industrial protein feeds to free  
CC amino acids, to degrade proteinaceous wastes, as additives in enzyme  
CC formulations used to treat malabsorption syndrome and for studying its  
CC biological role. Antibodies against Amp are used in immunohistochemical  
CC methods to study Amp distribution.

XX  
XX  
SQ Sequence 44453 BP; 13034 A; 10110 C; 10292 G; 11017 T; 0 other;

Query March 0.7%; Score 20; DB 20; Length 44453;  
Best Local Similarity 100.0%; Pred. No. 60;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1542 CCAAAAGCGAGAAATTGGCC 1661  
|||||  
Db 39033 CCAAAAGCGAGAAATTGGCC 39014

Search completed: October 3, 2003, 00:12:29  
Job time : 730 secs

0

KM autoimmune disorder; rheumatoid arthritis; multiple sclerosis; tumour;  
 KM autoimmune thyroiditis; allergic reaction; neurodegenerative disease;  
 KM Alzheimer's disease; Parkinson's disease; liver fibrosis; cancer; ulcer;  
 KM coagulation disorder; inflammatory disorder; Crohn's disease; incision;  
 KM tissue regeneration; wound healing; burn; haematopoiesis;  
 KM myeloid cell deficiency; lymphoid cell deficiency.  
 OS Homo sapiens.  
 XX  
 XX WO200177288-A2.  
 PN  
 XX  
 PD 18-OCT-2001.  
 XX  
 PF 29-MAR-2001; 2001WO-US10224.  
 PR  
 XX 06-APR-2000; 2000US-195582P.  
 XX  
 PA (GEMV ) GENETICS INST INC.  
 XX  
 PI Wong GG, Clark HF, Fechtel K, Agostino MJ, Howes SH, Resnick RJ;  
 PI Gulkrota K, Graham JR;  
 DR WPI; 2002-179321/23.  
 XX  
 PT Five hundred and ninety two polynucleotides derived from a variety of  
 PT human tissue sources which encode secreted proteins, useful for  
 PT treating immune deficiencies and disorders such as autoimmune disorders  
 PT  
 PS  
 XX Claim 1; Page 78; 372pp; English.  
 XX  
 CC The invention relates to 592 polynucleotides which have been derived from  
 CC a variety of human tissue sources and which encode novel secreted  
 CC proteins. The polynucleotides can be used as probes for the  
 CC identification and isolation of full length cDNA and genomic DNA. The  
 CC polynucleotides and proteins can also be used as nutritional supplements.  
 CC The proteins are useful in the treatment of various immune deficiencies  
 CC and disorders such as viral infections, bacterial infections, fungal  
 CC infections, autoimmune disorders (e.g. rheumatoid arthritis, multiple  
 CC sclerosis, autoimmune thyroiditis and diabetes) and allergic reactions  
 CC and conditions (e.g. asthma). They are also useful for treating  
 CC neurodegenerative diseases (e.g. Alzheimer's disease, Parkinson's  
 CC disease), liver fibrosis, coagulation disorders (e.g. haemophilia),  
 CC inflammatory disorders (e.g. Crohn's disease) and tumours. They are also  
 CC useful for tissue regeneration, for wound healing and in the treatment of  
 CC burns, incisions and ulcers. The proteins are also useful for regulating  
 CC haematopoiesis and for treating myeloid or lymphoid cell deficiencies.  
 CC Sequences ABK34863-ABK35454 represent polynucleotides of the invention.  
 CC  
 XX  
 SQ Sequence 2493 BP; 639 A; 520 C; 800 G; 533 T; 1 other;  
 Query Match 0.74; Score 20; DB 24; Length 2493;  
 Best Local Similarity 100.0%; Pred. No. 56;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2476 ATTCTGCGCCACCTGCTCC 2495  
 Db 564 ATTCTGCGCCACCTGCTCC 545  
 RESULT 43  
 ABL16061  
 ID ABL16061 standard; cDNA; 2706 BP.  
 XX  
 AC ABL16061;  
 XX  
 XX 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 42665.  
 XX  
 KM Drosophila; developmental biology; cell signalling; insecticide;  
 KM pharmaceutical; gene; ss.  
 XX  
 PA  
 XX

OS Drosophila melanogaster.  
 XX  
 XX WO200171042-A2.  
 PN  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US09231.  
 PR  
 XX 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) <sup>9</sup>Pe COR<sup>3</sup> NY.  
 XX  
 PI Venter JC, Adams M, Li FWD, Myers EW;  
 PI P-PSDB; ABB71958.  
 DR WPI; 2001-656860/75.  
 DR P-PSDB; ABB71958.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 PT  
 PS Claim 1; SEQ ID NO 42665; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (AB16176-AB16511), expressed DNA  
 CC sequences (AB101840-AB16175) and the encoded proteins  
 CC (AB57737-AB822072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 XX  
 SQ Sequence 2706 BP; 786 A; 730 C; 727 G; 463 T; 0 other;  
 Query Match 0.74; Score 20; DB 23; Length 2706;  
 Best Local Similarity 100.0%; Pred. No. 56;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1869 ACAGCAAGAAAAAAGGAG 1888  
 Db 1565 ACAGCAAGAAAAAAGGAG 1584  
 RESULT 44  
 ABL16060/c  
 ID ABL16060 standard; cDNA; 4757 BP.  
 XX  
 AC ABL16060;  
 XX  
 XX 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 42662.  
 XX  
 KM Drosophila; developmental biology; cell signalling; insecticide;  
 KM pharmaceutical; gene; ss.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US09231.  
 PR  
 XX 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PA

```

AC ABS99353;
XX 18-DEC-2002 (first entry)
XX
XX Enterococcus faecalis contig sequence #621.
DE
XX Computer readable medium; Enterococcus faecalis; microbe; growth;
XX pathogenicity; vaccine; resistance; Enterococcal infection; commercial;
XX therapeutic; industrial; fermenting; sugar source; metabolite; vaccine;
XX biocamp technology; antibacterial; modulator of nucleic acid expression;
XX contig; ds.
XX
XX Enterococcus faecalis.
OS
XX US2002120116-A1.
XX
XX 29-AUG-2002.
XX
XX 04-MAY-1998; 98US-0070927.
XX
XX 04-MAY-1998; 98US-0070927.
XX
XX 04-MAY-1998; 98US-0070927.
XX
XX (KUNSCH C A.
XX (DILLON P J.
XX (BARASH S.
XX
XX Kunsch CA, Dillon PJ, Barash S;
XX
XX WPI; 2002-750065/81.
XX
XX Computer readable medium having recorded on it a Enterococcus faecalis
XX PT nucleotide sequence useful for detecting diseases related to
XX PT Enterococcus infections in animals
XX
XX Claim 1; Page -; 119pp; English.
XX
XX The present invention relates to a new computer readable medium with an
XX CC Enterococcus faecalis nucleotide sequence. The invention is useful to
XX CC diagnose the presence of E.faecalis in a sample or determining the
XX CC presence of a specific microbe in a sample. The invention is also useful
XX CC for modulating the growth or pathogenicity of E.faecalis, in a vaccine
XX CC to confer resistance to Enterococcal infection, for commercial,
XX CC therapeutic and industrial purposes, and for fermenting a particular
XX CC sugar source or to produce a particular metabolite. The invention is
XX CC useful for detecting diseases related to Enterococcus infections in
XX CC animals, and for detecting E.faecalis using biocamp technology. The
XX CC present nucleic acid sequence represents an Enterococcus faecalis contig
XX CC DNA sequence of the invention.
XX CC Note: The sequence data for this patent did not form part of the
XX CC printed specification but was obtained in electronic format directly
XX CC from USPTO at http://seqdata.uspto.gov.
XX
XX SQ Sequence 1635 BP; 570 A; 298 C; 308 G; 448 T; 11 other;
XX
XX Query Match 0.7%; Score 20; DB 24; Length 1635;
XX Best Local Similarity 100.0%; Pred. No. 56;
XX Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 38 TGGCTGCTTCTTCAGATTCT 57
XX |||||||||||||||||||
XX DB 766 TGGCTGCTTCTTCAGATTCT 785
XX
XX RESULT 41
XX AAS90001
XX ID AAS90001 standard; cDNA; 1817 BP.
XX
XX AAS90001;
XX
XX 13-FEB-2002 (first entry)
XX
XX DNA encoding novel human diagnostic protein #25805.
XX

```

```

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US608611.
XX
XX 31-MAR-2000; 2000US-0540217.
XX 23-AUG-2000; 2000US-0649167.
XX
XX (HYSEQ INC.
XX
XX Drmanac ET, Lie C, Tang YT;
XX
XX WPI; 2001-639362/73.
XX P-PSCB; ABG25814.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity
XX
XX Claim 1; SEQ ID No 25805; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX CC and gene mapping, and in recombinant production of (II). The
XX CC polynucleotides are also used in diagnostics as expressed sequence tags
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques
XX CC to restore normal activity of (II) or to treat disease states involving
XX CC (II). (II) is useful for generating antibodies against it, detecting or
XX CC quantifying a polypeptide in tissue, as molecular weight markers and as
XX CC a food supplement. (II) and its binding partners are useful in medical
XX CC imaging of sites expressing (II). (I) and (II) are useful for treating
XX CC disorders involving aberrant protein expression or biological activity.
XX CC The polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC responsible for genetic disorders or other traits to assess biodiversity
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. AAS64197-AAS94564 represent novel human
XX CC diagnostic coding sequences of the invention.
XX CC Note: The sequence data for this patent did not appear in the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published\_pct\_sequences.
XX
XX SQ Sequence 1817 BP; 380 A; 492 C; 520 G; 425 T; 0 other;
XX
XX Query Match 0.7%; Score 20; DB 23; Length 1817;
XX Best Local Similarity 100.0%; Pred. No. 56;
XX Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 2476 ATTGCTGGGCACTGCTGCC 2495
XX |||||||||||||||||||
XX DB 287 ATTGCTGGGCACTGCTGCC 306
XX
XX RESULT 42
XX ABR34875/C
XX ID ABR34875 standard; cDNA; 2493 BP.
XX
XX ABR34875;
XX
XX 08-MAY-2002 (first entry)
XX
XX Human cDNA encoding secreted protein #13.
XX
XX Human; secreted protein; gene; ss; nutritional supplement; haemophilia;
XX KM viral infection; bacterial infection; fungal infection; diabetes; asthma;
XX

```

CY 38 TGCTGCTTCTCAGAAATCT 57  
Db 726 TGCTGCTTCTCAGAAATCT 745

RESULT 38  
AA553120  
ID AA553120 standard; DNA; 1299 BP.

AC AA553120;  
DT 13-FEB-2002 (first entry)

DE Enterococcus faecalis DNA for cellular proliferation protein #548.

KM Antisense; ds; prokaryotic cellular proliferation gene;  
KW antibiotic; antibacterial; drug design.

OS Enterococcus faecalis.

PN M0200170955-A2.

PD 27-SEP-2001.

PF 21-MAR-2001; 2001WO-US09180.

PR 21-MAR-2000; 2000US-191078P.

PR 23-MAY-2000; 2000US-206884P.

PR 26-MAY-2000; 2000US-207727P.

PR 23-OCT-2000; 2000US-242578P.

PR 27-NOV-2000; 2000US-253625P.

PR 22-DEC-2000; 2000US-257931P.

PR 16-FEB-2001; 2001US-269308P.

XX (ELITRA) ELITRA PHARM INC.

PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
PI Yamamoto RT, Xu HH;

DR WPI: 2001-611495/70.  
P-PSDB: AAU35261.

PS New polynucleotides for the identification and development of  
PT antibiotics, comprise sequences of antisense nucleic acids -  
XX Claim 27: Seq ID No 6757: 51np: English.

CC The invention relates to antisense inhibitors of genes essential to  
CC prokaryotic cellular proliferation, their use in identifying the  
CC genes, their use in the discovery of novel antibiotics, the essential  
CC genes themselves and the encoded proteins. The prokaryotes used are  
CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*  
CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The  
CC invention is also useful for the identification of potential new targets  
CC for antibiotic development. The antisense nucleic acid can also be used  
CC to identify proteins used in proliferation, to express these proteins,  
CC and to obtain antibodies capable of binding to the expressed proteins.  
CC The proteins can be used to screen compounds in rational drug discovery  
CC programmes. The antisense nucleic acid sequence is also useful to screen  
CC for homologous nucleic acids which are required for cell proliferation in  
CC a wide variety of organisms. The present sequence encodes an  
CC essential prokaryotic cellular proliferation protein.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC [http://wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences).

XX Sequence 1299 BP; 432 A; 250 C; 268 G; 349 T; 0 other;

Query Match 0.7%; Score 20; DB 23; Length 1299;  
Best Local Similarity 100.0%; Pred. No. 55;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 39 TGCTGCTTCTCAGAAATCT 57  
Db 726 TGCTGCTTCTCAGAAATCT 745

RESULT 39  
AA13558  
ID AA13558 standard; DNA; 1635 BP.

AC AA13558;  
DT 19-MAR-1999 (first entry)

DE Enterococcus faecalis genome contig SEQ ID NO:621.

KM Enterococcus faecalis; contig; detection; Enterococcal infection;  
KW vaccine; attenuation; computer readable medium; ds.

OS Enterococcus faecalis.

PN M03850555-A2.

PD 12-NOV-1998.

PF 04-MAY-1998; 98WO-US09985.

PR 14-NOV-1997; 97US-0066009.

PR 06-MAY-1997; 97US-0044031.

PR 16-MAY-1997; 97US-0046655.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Barash SC, Dillon PJ, Kunsch CA;

DR WPI: 1999-045171/04.

PT New isolated Enterococcus faecalis polynucleotides and polypeptides  
PT - used to develop products for the detection of Enterococcus and for  
PT use in vaccines for prevention or attenuation of Enterococcus  
PT infection.

XX Claim 1: Page 1866-1867; 2084pp: English.

CC A computer readable medium has been developed which has recorded on it  
CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.  
CC AA13558 to AA13599 represent these nucleotide sequences which are  
CC primary nucleotide sequences, also known as contigs. The computer-based  
CC system can identify fragments of the Enterococcus faecalis genome with  
CC commercial importance. The products can be used to detect the presence  
CC of Enterococcus faecalis in samples. They can also be used for  
CC diagnosing Enterococcal infection in an animal and monitoring  
CC progression of disease, and for identifying agents which can be used to  
CC modulate the growth or pathogenicity of Enterococcus faecalis, or  
CC another related organism, in vivo or in vitro. In particular the  
CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences  
CC can be used in vaccines to prevent or attenuate an Enterococcal  
CC infection.

XX Sequence 1635 BP; 570 A; 296 C; 308 G; 448 T; 11 other;

Query Match 0.7%; Score 20; DB 20; Length 1635;  
Best Local Similarity 100.0%; Pred. No. 56;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 38 TGCTGCTTCTCAGAAATCT 57  
Db 766 TGCTGCTTCTCAGAAATCT 785

RESULT 40  
ABS99353  
ID ABS99353 standard; DNA; 1635 BP.



PR 29-OCT-1999; 99US-0162142.

# Query Match

Best Local Similarity 100.0%; Score 20; DB 21; Length 638;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 746 TTTTACCCAAAAAGATATAA 765  
Db 22 TTTTACCCAAAAAGATATAA 41

## RESULT 36

AAF92514 ID AAF92514 standard; DNA; 954 BP.

AAF92514;

16-MAY-2001 (first entry)

Human T2R14 nucleotide sequence SEQ ID NO:27.

Human; rat; mouse; T2R; taste receptor; G-protein coupled receptor;

taste transduction G-protein coupled receptor; identification; tongue;

taste sensory neuron; taste cell; taste modulator; food;

taste signalling pathway; ds.

Homo sapiens.

MO200118050-A2.

15-MAR-2001.

08-SEP-2000; 2000MO-US24821.

10-SEP-1999; 99US-0393634.

22-FEB-2000; 2000US-0510332.

(REGC) UNIV CALIFORNIA.

(USSH) US DEPT HEALTH & HUMAN SERVICES.

Zucker CS, Adler JE, Ryba N, Mueller K, Hoon M;

WPI; 2001-211396/21.

P-PSDB; AAB87744.

Disclosure; Page 171-172; 249pp; English.

AAF92502 to AAF92572 represent nucleic acids which encode taste

transduction G-protein coupled receptors designated T2R proteins.

AAF87731 to AAB87824 represent T2R proteins, and AAB87825 to AAB87830

represent T2R family consensus sequences from the present invention.

The T2R proteins are taste modulators. The nucleic acids are useful as

probes for the identification of taste cells, as the nucleic acids are

specifically expressed in taste cells. They also serve as tools for the

generation of taste topographic maps that elucidate the relationship

between the taste cells of the tongue and taste sensory neurons leading

to taste centres in the brain. The taste modulators are useful for

pharmacological and genetic modulation of taste signalling pathways.

Modulatory compounds comprising T2R proteins can therefore be used in

food and pharmaceutical industries to customise taste, for e.g. to

decrease the bitter taste of food or drugs.

Sequence 954 BP; 249 A; 183 C; 194 G; 328 T; 0 other;

Query Match 0.7%; Score 20; DB 22; Length 954;

Best Local Similarity 100.0%; Pred. No. 55;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 29 TGGTTTGGTGTGCTCTT 48

Db 386 TGGTTTGGTGTGCTCTT 405

AA551375 ID AA551375 standard; DNA; 1287 BP.

AA551375;

13-FEB-2002 (first entry)

Enterococcus faecalis DNA for cellular proliferation protein #152.

Antisense; ds; prokaryotic cellular proliferation gene;

antibiotic; antibacterial; drug design.

Enterococcus faecalis.

WO200170955-A2.

21-MAR-2001; 2001MO-US09180.

21-MAR-2001; 2000US-191078P.

23-MAY-2000; 2000US-206848P.

26-MAY-2000; 2000US-207727P.

23-OCT-2000; 2000US-242578P.

27-NOV-2000; 2000US-253625P.

22-DEC-2000; 2000US-257931P.

16-FEB-2001; 2001US-269308P.

(ELITRA) ELITRA PHARM INC.

Haselbeck R, Ohlser KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

Yamamoto RT, Xu HH;

WPI; 2001-611495/70.

P-PSDB; AAU33516.

New polynucleotides for the identification and development of

antibiotics, comprise sequences of antisense nucleic acids.

Claim 27; Seq ID No 3957; 51pp; English.

The invention relates to antisense inhibitors of genes essential to

prokaryotic cellular proliferation, their use in identifying the

genes, their use in the discovery of novel antibiotics, the essential

genes themselves and the encoded proteins. The prokaryotes used are

Escherichia coli, Streptococcus aureus, Salmonella typhi, Klebsiella

pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The

invention is also useful for the identification of potential new targets

for antibiotic development. The antisense nucleic acids can also be used

to identify proteins used in proliferation, to express these proteins,

and to obtain antibodies capable of binding to the expressed proteins.

The proteins can be used to screen compounds in rational drug discovery

programmes. The antisense nucleic acid sequence is also useful to screen

for homologous nucleic acids which are required for cell proliferation in

a wide variety of organisms. The present sequence encodes an

essential prokaryotic cellular proliferation protein.

Note: the sequence data for this patent did not form part

of the printed specification, but was obtained in electronic

format directly from WIPO at

ftp.wipo.int/pub/published\_pct\_sequences.

Sequence 1287 BP; 422 A; 249 C; 268 G; 348 T; 0 other;

Query Match 0.7%; Score 20; DB 23; Length 1287;

Best Local Similarity 100.0%; Pred. No. 55;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140891.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.

PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149375.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 26-SEP-1999; 99US-0156459.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0158293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161921.  
PR 28-OCT-1999; 99US-0161993.

PS Example 4, Page 213: 378bp; English.

XX The invention relates to isolated lung tumour-specific proteins and  
 CC their corresponding cDNA molecules. Lung tumour-specific proteins and  
 CC their antigen-presenting cells are useful for stimulating and/or  
 CC expanding T cells specific for a tumour protein, and for inhibiting  
 CC the development of cancer. The invention also relates to a composition  
 CC useful for stimulating an immune response, and for treating cancer. The  
 CC lung tumour specific oligonucleotide is useful in gene therapy and for  
 CC diagnosis, detection and treatment of lung cancer. The present sequence  
 CC is a cDNA encoding human lung tumour-specific protein.

XX Sequence 548 BP, 180 A; 97 C; 180 G; 91 T; 0 other;

SO Query Match 0.7%; Score 20; DB 23; Length 548;  
 Best Local Similarity 100.0%; Pred. No. 54;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2476 ATTCTGGGCGACCTGGTCC 2495  
 Db 543 ATTCTGGGCGACCTGGTCC 524

RESULT 34  
 ABL83527  
 ID ABL83527 standard; cDNA; 577 BP.

XX ABL83527;  
 AC 17-MAY-2002 (first entry)  
 DT  
 XX Human ovarian cancer related cDNA clone SEQ ID NO:6505.  
 DE  
 XX Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.  
 KM  
 XX Homo sapiens.  
 OS  
 XX WO200192581-A2.  
 PN  
 XX 06-DEC-2001.  
 PD  
 XX 29-MAY-2001; 2001WO-US17756.  
 PF  
 XX 26-MAY-2000; 2000US-207484P.  
 PR  
 XX (CORI-) CORIXA CORP.  
 PA  
 XX A1gate PA, Harlocker SL, Jones R;  
 PI  
 XX WPI; 2002-122075/16.  
 DR  
 XX Composition for therapy and diagnosis of ovarian cancer comprising  
 PT polypeptide of a ovarian tumor polypeptide; polynucleotide encoding  
 PT polypeptide, antibody specific to polypeptide or T cell expressing  
 PT polypeptide -  
 PT  
 XX Claim 1; SEQ ID 6505; 489bp; English.

PS The present invention describes a composition (I) comprising: carriers  
 CC and immunostimulants; and a polypeptide (III) of a ovarian tumor  
 CC polypeptide encoded by a polynucleotide (III) having a cDNA sequence  
 CC (SI) from the 10912 nucleotide sequences as given in ABL77023 to  
 CC ABL7934, (III) encoding (II) having a sequence (S2), a T cell  
 CC population of (II), or antigen presenting cells that express (II).  
 CC (I) has cytostatic activity. An oligonucleotide (IV) that hybridizes to  
 CC (SI) can be used for detecting ovarian cancer in a patient's biological  
 CC sample preferably serum or ovarian tissue. The method comprises  
 CC contacting a biological sample from a patient with (IV), detecting the  
 CC amount of polynucleotide hybridizing to (IV) and comparing the amount to  
 CC a predetermined cutoff value and thereby detecting ovarian cancer in the  
 CC patient, where the amount of polynucleotide hybridizing to (IV) is  
 CC detected preferably by polymerase chain reaction (PCR). (I) comprising  
 CC (III) and/or (II) is useful for stimulating and/or expanding T cells

CC specific for an ovarian tumor protein comprising contacting T cells  
 CC with (III) or (II). (III) is useful in design and preparation of  
 CC ribozyme molecules for inhibiting expression of the tumor polypeptides  
 CC and proteins in tumor cells; and to isolate a full length gene from a  
 CC suitable library e.g., a tumor cDNA library using well known  
 CC techniques.

XX Sequence 577 BP, 152 A; 121 C; 117 G; 187 T; 0 other;

SO Query Match 0.7%; Score 20; DB 24; Length 577;  
 Best Local Similarity 100.0%; Pred. No. 54;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2505 AGAGGTCGACAGAGATAC 2524  
 Db 345 AGAGGTCGACAGAGATAC 364

RESULT 35  
 AAC38427  
 ID AAC38427 standard; DNA; 639 BP.

XX AAC38427;  
 AC 17-OCT-2000 (first entry)  
 DT  
 XX Arabidopsis thaliana DNA fragment SEQ ID NO: 20930.  
 DE  
 XX Hybridisation assay; genetic mapping; gene expression control;  
 KW protein identification; signal transduction pathway;  
 KW metabolic pathway; promoter; termination sequence; ss.  
 XX  
 OS Arabidopsis thaliana.  
 PN  
 XX EP1033405-A2.  
 PD  
 XX 06-SEP-2000.  
 PF  
 XX 25-FEB-2000; 2000EP-0301439.  
 PR  
 XX 25-FEB-1999; 99US-0121825.  
 PR 05-MAR-1999; 99US-0123180.  
 PR 09-MAR-1999; 99US-0123548.  
 PR 23-MAR-1999; 99US-0125788.  
 PR 29-MAR-1999; 99US-0126264.  
 PR 29-MAR-1999; 99US-0126785.  
 PR 01-APR-1999; 99US-0127462.  
 PR 06-APR-1999; 99US-0128234.  
 PR 08-APR-1999; 99US-0128714.  
 PR 16-APR-1999; 99US-0128945.  
 PR 19-APR-1999; 99US-0130077.  
 PR 21-APR-1999; 99US-0130449.  
 PR 23-APR-1999; 99US-0130510.  
 PR 23-APR-1999; 99US-0130891.  
 PR 28-APR-1999; 99US-0131449.  
 PR 30-APR-1999; 99US-0132048.  
 PR 30-APR-1999; 99US-0132407.  
 PR 04-MAY-1999; 99US-0132484.  
 PR 05-MAY-1999; 99US-0132485.  
 PR 06-MAY-1999; 99US-0132486.  
 PR 06-MAY-1999; 99US-0132487.  
 PR 07-MAY-1999; 99US-0132863.  
 PR 11-MAY-1999; 99US-0134256.  
 PR 14-MAY-1999; 99US-0134218.  
 PR 14-MAY-1999; 99US-0134219.  
 PR 14-MAY-1999; 99US-0134221.  
 PR 14-MAY-1999; 99US-0134370.  
 PR 18-MAY-1999; 99US-0134768.  
 PR 19-MAY-1999; 99US-0134941.  
 PR 20-MAY-1999; 99US-0135124.  
 PR 21-MAY-1999; 99US-0135353.  
 PR 24-MAY-1999; 99US-0135629.  
 PR 25-MAY-1999; 99US-0136021.

PF 26-JAN-1999; 99MO-US01642.  
 XX  
 PR 22-DEC-1998; 98US-0219245.  
 PR 28-JAN-1998; 98US-0015022.  
 PR 28-JAN-1998; 98US-0015029.  
 PR 18-MAR-1998; 98US-0040828.  
 PR 18-MAR-1998; 98US-0040831.  
 PR 23-JUL-1998; 98US-0122191.  
 PR 23-JUL-1998; 98US-0122192.  
 XX  
 PA (CORI-) CORIXA CORP.  
 PI Fradakis TN, Lodes MJ, Mohamath R, Reed SG;  
 DR WPI: 1999-479187/40.  
 DR P-PSDB; AAY29561.  
 XX  
 PT Lung tumour specific polynucleotides for inhibiting the development  
 of lung cancer  
 XX  
 PS Claim 13; Page 144; 171pp; English.  
 CC The present invention describes lung tumour specific polynucleotides  
 CC and tumour antigens. AA207144 to AA207246 and AA208301 to AA208325  
 CC represent specifically claimed polynucleotides, and AAY29486 to AAY29571  
 CC represent amino acid sequences from the present invention. The lung  
 CC tumour specific polynucleotides and polypeptides can be used in  
 CC pharmaceutical compositions and vaccines to inhibit the development of  
 CC lung cancer. They can also be used to detect lung cancer in a patient.  
 CC Probes and antibodies derived from the lung tumour sequences are useful  
 CC in detection of lung cancer.  
 CC  
 SQ Sequence 548 BP; 180 A; 97 C; 180 G; 91 T; 0 other;  
 OY Query Match 0.7%; Score 20; DB 20; Length 548;  
 Best Local Similarity 100.0%; Pred. No. 54;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 2476 ATTCCTGGGCGACCTGGTCC 2495  
 543 ATTCCTGGGCGACCTGGTCC 524  
 RESULT 32  
 AAC79168/c  
 ID AAC79168 standard; CDNA; 548 BP.  
 AC AAC79168;  
 XX  
 DT 05-FEB-2001 (first entry)  
 XX  
 DE Human lung tumour-specific cDNA #121.  
 XX  
 KW Lung tumour protein; lung cancer; cytostatic; vaccine; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200060077-A2.  
 XX  
 PD 12-OCT-2000.  
 XX  
 PE 30-MAR-2000; 2000MO-US08560.  
 XX  
 PR 02-APR-1999; 99US-0285323.  
 PR 09-AUG-1999; 99US-0370838.  
 PR 30-DEC-1999; 99US-0476235.  
 PR 03-MAR-2000; 2000US-0518809.  
 XX  
 PA (CORI-) CORIXA CORP.  
 PI Reed SG, Lodes MJ, Mohamath R, Secrist H;  
 XX  
 DR WPI: 2000-638466/61.

DR P-PSDB; AAB44485.  
 XX  
 PT Novel lung tumor polypeptides and polynucleotides, useful for  
 PT detecting, monitoring or treating cancer, especially lung cancer -  
 XX  
 PS Claim 1; Page 159; 243pp; English.  
 CC The present sequence is given in a specification relating to compounds  
 CC for therapy and diagnosis of lung cancer. Polypeptides comprising at  
 CC least an immunogenic part of a lung tumour protein are disclosed.  
 CC The polypeptides are useful for inhibiting the development of cancer,  
 CC especially lung cancer. Samples of T cells expressing the polypeptides  
 CC may be used to inhibit the development of cancer. The polypeptides are  
 CC also useful for detecting and monitoring the progression of cancer,  
 CC especially lung cancer.  
 CC  
 SQ Sequence 548 BP; 180 A; 97 C; 180 G; 91 T; 0 other;  
 OY Query Match 0.7%; Score 20; DB 21; Length 548;  
 Best Local Similarity 100.0%; Pred. No. 54;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 2476 ATTCCTGGGCGACCTGGTCC 2495  
 543 ATTCCTGGGCGACCTGGTCC 524  
 RESULT 33  
 AAD23243/c  
 ID AAD23243 standard; CDNA; 548 BP.  
 AC AAD23243;  
 XX  
 DT 26-FEB-2002 (first entry)  
 XX  
 DE Human lung tumour-specific protein SAL-68 cDNA.  
 XX  
 KW Human; lung tumour protein; immunostimulant; cytostatic; gene therapy;  
 KW antisense-therapy; vaccine; immune response; lung cancer; SAL-68; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT 17..547  
 FT CDS /tag= a  
 FT /product= "Human SAL-68 protein"  
 FT /note= "CDS does not include stop codon"  
 FT /partial  
 PN WO200172295-A2.  
 XX  
 PD 04-OCT-2001.  
 XX  
 PE 28-MAR-2001; 2001MO-US09991.  
 XX  
 PR 29-MAR-2000; 2000US-0538037.  
 PR 05-JUN-2000; 2000US-0588937.  
 PR 18-AUG-2000; 2000US-0640878.  
 PR 22-SEP-2000; 2000US-234517P.  
 PR 01-NOV-2000; 2000US-0704512.  
 PR 14-DEC-2000; 2000US-0738973.  
 XX  
 PA (CORI-) CORIXA CORP.  
 PI Reed SG, Lodes MJ, Mohamath R, Secrist H, Benson DR, Indrias CY;  
 PI Henderson RA, Fling SP, Algate PA, Ellitt M, Mannion J, Kalos MD;  
 XX  
 DR WPI: 2001-639201/73.  
 DR P-PSDB; AAE13826.  
 XX  
 PT New human lung-specific polynucleotides and polypeptides for the  
 PT diagnosis and treatment of disease e.g. lung cancer -

PS Claim 1; SEQ ID No 5973; 658bp; English.

XX

CC The invention relates to a single exon nucleic acid probe (SENP) (:) for

CC measuring human gene expression in a sample derived from human adult

CC liver, comprising one of 13109 defined nucleotide sequences given in the

CC specification (or complements/ fragments). The probe hybridises at high

CC stringency to a nucleic acid molecule expressed in the human adult

CC liver. (i) may be used for predicting, measuring and displaying gene

CC expression in samples derived from human adult liver. The genes

CC identified may be involved in genetic liver diseases such as cirrhosis,

CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which

CC is associated with coronary heart disease. ABS5011-ABS51005 represent

CC human liver single exon nucleic acid probes of the invention.

CC Note: The sequence information for this patent does not appear in the

CC printed specification but was obtained in electronic format directly

CC from Wipo.int/ftp/pub/published\_pct\_sequences.

XX

SC Sequence 479 BP; 169 A; 82 C; 105 G; 123 T; 0 other;

Query Match 0.7%; Score 20; DB 23; Length 479;

Best Local Similarity 100.0%; Pred. No. 54;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 TGGTTTGGTCTGCTCTT 48

Db 386 TGGTTTGGTCTGCTCTT 367

RESULT 30

ABS06054/C

ID ABS06054 standard; DNA; 479 BP.

XX

AC ABS06054;

XX

DT 19-AUG-2002 (first entry)

XX

DE Human genome-derived single exon probe from lung SEQ ID No 6045.

XX

KM Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;

KM chronic obstructive pulmonary disease; interstitial lung disease;

KM familial idiopathic pulmonary fibrosis; neurofibromatosis;

KM tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;

KM Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;

KM pulmonary histiocytosis; lymphangioleiomyomatosis; Karsenger syndrome;

KM pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;

KM primary ciliary dyskinesia; pulmonary hypertension;

KM Fyalline membrane disease.

XX

OS Homo sapiens.

XX

PN WO200186003-A2.

XX

PD 15-NOV-2001.

XX

XX 30-JAN-2001; 2001WO-US00665.

PF

XX 04-FEB-2000; 2000US-180312P.

PR 26-MAY-2000; 2000US-207456P.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-234687P.

PR 27-SEP-2000; 2000US-236359P.

PR 04-OCT-2000; 2000GB-0024263.

XX

PA (MOLE-) MOLECULAR DYNAMICS INC.

XX

PI Penn SQ, Hanzel DK, Chen W, Rank DR;

XX

DR WPI; 2002-114183/15.

XX

PT Spatially-addressable set of single exon nucleic acid probes, used to

PT measure gene expression in human lung samples -

XX

PS Claim 1; SEQ ID No 6045; 634bp; English.

XX

CC The invention relates to a spatially-addressable set of single exon

CC nucleic acid probes for measuring gene expression in a sample derived

CC from human lung comprising single exon nucleic acid probes having one of

CC 12614 nucleic acid sequences mentioned in the specification, or their

CC complements or the 12387 open reading frames derived from the 12614

CC probes. Also included are a microarray comprising the novel set of

CC probes; the novel set of probes which hybridise at high stringency to a

CC nucleic acid expressed in the human lung; measuring gene expression in a

CC sample derived from human lung, comprising (a) contacting the array with

CC a collection of detectably labeled nucleic acids derived from human lung

CC mRNA, and (b) measuring the label detectably bound to each probe of

CC the array; identifying exons in a eukaryotic genome, comprising

CC (a) algorithmically predicting at least one exon from genomic sequences

CC of the eukaryote; and (b) detecting specific hybridisation of detectably

CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,

CC having a fragment identical to the predicted exon. The probe is included

CC in the above mentioned microarray; assigning exons to a single gene,

CC comprising (a) identifying exons from genomic sequence by the method

CC above and (b) measuring the expression of each of the exons in several

CC tissues and/or cell types using hybridisation to a single exon

CC microarrays having a probe with the exon. Where a common pattern of

CC expression of the exons in the tissues and/or cell types indicates that

CC the exons should be assigned to a single gene, a peptide comprising one

CC of 12011 sequences, mentioned in the specification, or encoded by the

CC probes/open reading frames (ORF). The probes are used for gene

CC expression analysis, and for identifying exons in a gene, particularly

CC using human lung derived mRNA and for the study of lung diseases

CC such as asthma, lung cancer, chronic obstructive pulmonary disease

CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary

CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,

CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary

CC haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,

CC pulmonary alveolar proteinosis, Karsenger syndrome, fibrocystic

CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension

CC and hyaline membrane disease. The present sequence is a single exon

CC probe of the invention.

CC Note: The sequence data for this patent did not form part

CC of the printed specification, but was obtained in electronic

CC format directly from Wipo.int/ftp/pub/published\_pct\_sequences.

XX

SC Sequence 479 BP; 169 A; 82 C; 105 G; 123 T; 0 other;

Query Match 0.7%; Score 20; DB 24; Length 479;

Best Local Similarity 100.0%; Pred. No. 54;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 TGGTTTGGTCTGCTCTT 48

Db 386 TGGTTTGGTCTGCTCTT 367

RESULT 31

AA208316/C

ID AA208316 standard; CDNA; 548 BP.

XX

AC AA208316;

XX

DT 13-OCT-1999 (first entry)

XX

DE Human lung tumour protein SAT-68 5' CDNA sequence.

XX

KM Human; lung tumour protein; therapy; diagnosis; lung cancer; vaccine;

KM immunotherapy; detection; inhibition; se.

XX

OS Homo sapiens.

XX

PN WO9938973-A2.

XX

PD 05-AUG-1999.

XX

```

DE Human bone marrow expressed single exon probe SEQ ID NO: 5859.
XX
XX Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukemia; lymphoma; myeloma; ss.
XX
OS Homo sapiens.
XX
XX MO200157276-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00668.
XX
XX 04-FEB-2000; 2000US-0180312.
XX
XX 26-MAY-2000; 2000US-0207456.
XX
XX 30-JUN-2000; 2000US-0608408.
XX
XX 03-AUG-2000; 2000US-0632366.
XX
XX 21-SEP-2000; 2000US-0234687.
XX
XX 27-SEP-2000; 2000US-0236359.
XX
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488900/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human bone marrow.
XX
XX Example 4; SEQ ID NO: 5859; 658bp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukemia and myeloma. The present sequence is one of
CC the probes of the invention.
XX
XX Sequence 479 BP; 169 A; 82 C; 105 G; 123 T; 0 other;
SQ
Query Match 0.7%; Score 20; DB 22; Length 479;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 29 TGGTTTGGTCTCTCTT 48
DB 386 TGGTTTGGTCTCTCTT 367

```

```

PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488997/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta.
XX
XX Claim 25; SEQ ID No 5885; 654bp; English.
XX
XX The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
XX Sequence 479 BP; 169 A; 82 C; 105 G; 123 T; 0 other;
SQ
Query Match 0.7%; Score 20; DB 22; Length 479;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 29 TGGTTTGGTCTCTCTT 48
DB 386 TGGTTTGGTCTCTCTT 367

```

```

RESULT 29
ABS30983/C
ID ABS30983 standard; DNA; 479 BP.
XX
XX ABS30983;
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00664.
XX
XX 04-FEB-2000; 2000US-0180312.
XX
XX 26-MAY-2000; 2000US-0207456.
XX
XX 30-JUN-2000; 2000US-0608408.
XX
XX 03-AUG-2000; 2000US-0632366.
XX
XX 21-SEP-2000; 2000US-0234687.
XX
XX 27-SEP-2000; 2000US-0236359.
XX
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488998/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human adult liver.
XX

```

CC with increased tolerance to abiotic stress. The present sequence is that  
 CC of an Arabidopsis thaliana stress regulated gene (AB212196-AB217574) used  
 CC in methods of the invention.  
 CC Note: The sequence data for this patent is not represented in the printed  
 CC specification but is based on sequence information supplied to Derwent by  
 CC the European Patent Office.  
 XX

Sequence 393 BP; 133 A; 81 C; 92 G; 87 T; 0 other;

Query Match 0.7%; Score 20; DB 24; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 54;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 746 TTTCCACCAAAAAGATATAA 765  
 Db 33 TTTCCACCAAAAAGATATAA 52

RESULT 25

ABAS7633/C  
 ID ABAS7633 standard; DNA; 479 BP.

XX ABAS7633;

DT 01-FEB-2002 (first entry)

DE Human foetal liver single exon nucleic acid probe #5938.

KM Human; foetal liver; gene expression; single exon nucleic acid probe; ss.

OS Homo sapiens.

PN WO200157277-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00669.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-483447/52.

PT Human genome-derived single exon nucleic acid probes useful for

PT analyzing gene expression in human foetal liver -

PS Claim 1; SEQ ID NO 5938; 639bp + sequence listing; English.

CC The invention relates to a single exon nucleic acid probe for

CC measuring human gene expression in a sample derived from human foetal

CC liver. The single exon nucleic acid probes may be used for predicting,

CC measuring and displaying gene expression in samples derived from human

CC foetal liver. The present sequence is a single exon nucleic acid

CC probe of the invention.

CC Note: The sequence data for this patent did not form part of the

CC printed specification, but was obtained in electronic format directly

CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 479 BP; 169 A; 82 C; 105 G; 123 T; 0 other;

XX Query Match 0.7%; Score 20; DB 22; Length 479;

XX Best Local Similarity 100.0%; Pred. No. 54;

XX Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 29 TGGTTTGGTGGCTGCTTCTT 48  
 Db 396 TGGTTTGGTGGCTGCTTCTT 367

RESULT 26

AAK05687/C  
 ID AAK05687 standard; DNA; 479 BP.

XX AAK05687;

DT 05-NOV-2001 (first entry)

DE Human brain expressed single exon probe S50 ID NO: 5678.

KM Human; brain expressed exon; gene expression analysis; probe;

KM microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;

OS Homo sapiens.

PN WO200157275-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00667.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-483446/52.

PT Single exon nucleic acid probes for analyzing gene expression in human

PT brains -

PS Example 4; SEQ ID NO: 5678; 650bp + Sequence Listing; English.

CC The present invention provides a number of single exon nucleic acid

CC probes which are derived from genomic sequences expressed in the human

CC brain. They can be used to measure gene expression in brain cell samples,

CC which may enable the diagnosis and improved treatment of nervous system

CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,

CC epilepsy and cancers. The present sequence is one of the probes of the

CC invention.

XX Sequence 479 BP; 169 A; 82 C; 105 G; 123 T; 0 other;

XX Query Match 0.7%; Score 20; DB 22; Length 479;

XX Best Local Similarity 100.0%; Pred. No. 54;

XX Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX ID AAK31302 standard; DNA; 479 BP.

XX AAK31302;

XX AAK31302;

XX 06-NOV-2001 (first entry)

```

PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145293.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 22-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 28-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157665.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.

```

```

PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

```

Query Match 0.7%; Score 20; D3 21; Length 393;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 20; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 746 TTTCACCAAAAGATATAA 765
Db 33 TTTCACCAAAAGATATAA 52

```

```

RESULT 24
ABZ14620

```

```

ID ABZ.4620 standard; DNA; 393 BP.

```

```

XX ABZ.4620;

```

```

DT 21-JAN-2003 (first entry)

```

```

XX Arabidopsis thaliana stress regulated gene SEQ ID NO 2425.

```

```

XX Arabidopsis thaliana; plant; gene; stress; transgenic; ds.

```

```

XX Arabidopsis thaliana.

```

```

XX WO200216655-A2.

```

```

XX 28-FEB-2002.

```

```

PF 24-AUG-2001; 2001WO-US26685.

```

```

XX 24-AUG-2000; 2000US-227866P.

```

```

PR 26-JAN-2001; 2001US-264647P.

```

```

PR 22-JUN-2001; 2001US-300111P.

```

```

XX (SCRI ) SCRIPPS RES INST.

```

```

PA (SYGN ) SYNGENTA PARTICIPATIONS AG.

```

```

XX Harper JF, Kreps J, Wang X, Zhu T;

```

```

PI WPI: 2002-304127/34.

```

```

DR Identifying a stress condition to which a plant cell has been exposed

```

```

PT and producing plants with increased tolerance to these abiotic stresses

```

```

XX -

```

```

PT Claim 144; SEQ ID NO 2425; 577bp + Sequence Listing; English.

```

```

XX The invention relates to identifying a stress condition to which a plant

```

```

CC cell has been exposed, comprising:

```

```

CC (a) contacting nucleic acid representative of expressed polynucleotides

```

```

CC in the plant cell with an array or probes representative of the plant

```

```

CC cell genome; and

```

```

CC (b) detecting a profile of expressed polynucleotides in the plant cell

```

```

CC characteristic of a stress response; the method is useful in the

```

```

CC production of transgenic plants, cells and seeds and in producing plants

```



PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating cancers and  
PT metastasis -  
XX  
PS Disclosure; SEQ ID NO 35958; 3071bp + Sequence listing; English.  
XX  
CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)  
CC amino acid sequences given in AAK62170 to AAK91921. (I) have cytostatic  
CC activity, and can be used in gene therapy and vaccine production. (I)  
CC proteins and polynucleotides may be used in the prevention, diagnosis and  
CC treatment of diseases associated with inappropriate (I) expression. For  
CC example, they may be used to treat disorders associated with decreased  
CC expression by rectifying mutations or deletions in a patient's genome  
CC that affect the activity of (I) by expressing inactive proteins or to  
CC supplement the patient's own production of (I). Additionally, (I)  
CC polynucleotides may be used to produce the secreted (I), by inserting  
CC the nucleic acids into a host cell and culturing the cell to express the  
CC protein. (I) proteins and polynucleotides may be used to prevent,  
CC diagnose and treat immune/hematopoietic-related diseases, especially  
CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703  
CC to AAK67694 represent human immune/hematopoietic antigen genomic  
CC sequences from the present invention. AAK54942 to AAK54950 and AAK62169  
CC represent sequences used in the exemplification of the present invention.  
XX  
SQ Sequence 2154 BP; 638 A; 443 C; 480 G; 593 T; 0 other;  
Query Match 0.8%; Score 22; DB 22; Length 2154;  
Best Local Similarity 100.0%; Pred. No. 6.1;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1178 CTGGAAGAGAAAGAACTGTTCA 1199  
|||  
DB 268 CTGGAAGAGAAAGAACTGTTCA 289  
RESULT 23  
AAC43283  
ID AAC43283 standard; DNA; 393 BP.  
XX  
AC AAC43283;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 38698.  
XX  
KW Hybridisation assay; genetic mapping; gene expression; control;  
KW protein identification; signal transduction pathway;  
KW metabolic pathway; promoter; termination sequence; ss.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EPI033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0128845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134758.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139464.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 24-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140662.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142355.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144088.  
PR 16-JUL-1999; 99US-0144088.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145066.

[illegible]

XX	05-SEP-2000;	2000US-229747P.
PR	05-SEP-2000;	2000US-229748P.
PR	05-SEP-2000;	2000US-229749P.
PR	05-SEP-2000;	2000US-229750P.
PR	05-SEP-2000;	2000US-229751P.
PR	05-SEP-2000;	2000US-230583P.
PR	06-SEP-2000;	2000US-230505P.
PR	06-SEP-2000;	2000US-230514P.
PR	06-SEP-2000;	2000US-230515P.
PR	06-SEP-2000;	2000US-230517P.
PR	06-SEP-2000;	2000US-230518P.
PR	06-SEP-2000;	2000US-230519P.
PR	06-SEP-2000;	2000US-230595P.
PR	06-SEP-2000;	2000US-230597P.
PR	06-SEP-2000;	2000US-230598P.
PR	06-SEP-2000;	2000US-230599P.
PR	06-SEP-2000;	2000US-230610P.
PR	06-SEP-2000;	2000US-230665P.
PR	06-SEP-2000;	2000US-230988P.
PR	07-SEP-2000;	2000US-230951P.
PR	07-SEP-2000;	2000US-231163P.
PR	07-SEP-2000;	2000US-231167P.
XX		
XX	(INCY-)	INCYTE GENOMICS INC.

PI Stuart J., Lincoln SE, Altus CM, Dufour GE, Chaturp MS, Hillman JL,  
PI Jones AL, yu yn SE, Rietzen RJ, Gietzen D, Liu TF, Yap PE, Dahl CR;  
PI Moniluyama MG, Bradley DL, Rohaag SD, Harris B, Roseberry AM;  
PI Gerstlin EH, Petralia CH, David MH, Paner SR, Flores V, Daffoo A;  
PI Marwaha R, Chen AJ, Chang SC, Au AP, Inman RR;  
XX  
XX  
WP1: 2002-383054/41.  
DR  
DR P-ESDB: ABG60115.

PT An isolated polynucleotide useful in diagnostics and therapeutics -  
XX  
PS Claim 1; Page 487-488; 686pp; English.

CC The invention relates to human diagnostic and therapeutic (dntp)  
CC polynucleotides and their associated polypeptides (dntp polypeptides).  
CC The sequences of the invention are used in the treatment and diagnosis of  
CC cell proliferative disorders (e.g. atherosclerosis, cirrhosis) cancers  
CC (e.g. tumours of the adrenal gland, bone, bone marrow, brain, breast,  
CC cervix, kidney, lung, ovary, pancreas, prostate, skin, spleen, testis or  
CC thymus), autoimmune/inflammatory disorders (e.g. asthma, bronchitis,  
CC psoriasis, osteoporosis), viral infections, bacterial infections, fungal  
CC infections, parasitic infections, developmental disorders (e.g. anaemia,  
CC epilepsy), seizure disorders (e.g. cerebral palsy, spina bifida),  
CC endocrine disorders (e.g. thrombosis, aneurysm), metabolic disorders  
CC (e.g. obesity, diabetes), neurological disorders (e.g. stroke,  
CC amyotrophic lateral sclerosis, multiple sclerosis), gastrointestinal  
CC disorders (e.g. ulcerative colitis, lysinuria) and transport disorders  
CC (e.g. myotonic dystrophy, catatonias, periphrax and neuropathy). Sequences  
CC ABK71535-ABK71809 represent human dntp polynucleotides of the invention.  
XX Sequence 1101 BP, 334 A, 191 C, 203 G, 373 T, 0 other;

Query Match 2.2%; Score 63; DB 24; Length 1101;  
 Best Local Similarity 100.0%; Pred. No. 1e-19;  
 Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	2195	AGAGAGTGAAGGGAGAACTGGTGCACAGGTGCATTGGATCAAAAGAGAA <sup>TT</sup> CTG	2254
Db	140	AGAGAGTGAAGGGAGAACTGGTGCACAGGTGCATTGGATCAAAAGAGAA <sup>TT</sup> CTG	159
CY	2255	GGG 2257	
Db	200	GGG 202	

RESULT 21  
AAI36639/C

ID AAT36639 standard; ENA; 466 BP.

AC AA136639

DT :7-OCT-2001 (first entry)

DE Probe #5325 used to measure gene expression in human placenta sample.

KW Probe; microarray; human; placenta; antenatal diagnosis;

KW genetic disorder; ss.

OS Homo sapiens

PN W02003157272-A2

PD 09-ATG-2001.

PF 33-JAN-2001; 2001WO-US00663.

PR 04-FEB-2000; 2000US-0180312.

PR 30-JUN-2000; 2000JS-0608408.

PR 21-SEP-2000; 2000US-0234687

PR 04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC  
2A  
22

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-488897/53.

PT Human genome-derived single exon nucleic acid probes useful for

2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
841

100

CC The present sequence is one such probe. The probes are useful for

CC expression in samples derived from human placenta. The probes are useful

XX

Best Local Similarity 100.0%; Pred. No. 0.0375;

.....

```
Cy      2573 GACGTCACAGTGTACAGAGATTAAAGG 2600
          |||...|||
Db      466 GACGTCACAGTGTACAGAGATTAAAGG 499
```

RESULT 22  
AAK81046

XX

XX

✕✕

XX  
0  
A  
L  
C  
A  
A

cytostatic; gene therapy; vaccine; metastasis; ds.

OS Homo sapiens

PN WC2C0157182-A2

PD 09-AJG-2001.

```

OS Homo sapiens.
XX
XX BPI033401-A2.
XX
XX 06-SEP-2000.
XX
XX 21-FEB-2000; 2000EP-0200610.
XX
XX 26-FEB-1999; 99JS-0122487.
XX
XX (GEST ) GENSET.
XX
XX Dumas Milne Edwards J, Duclert A, Giordano J;
XX
XX WPI; 2000-500381/45.
XX
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
XX diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
XX Claim 1; SEQ ID 33926; 71bp + CD-ROM; English.
XX
XX The present sequence is one of a large number of 5' ESTs derived from
XX mRNAs encoding secreted proteins. No ORF has yet been conclusively
XX identified within the present sequence. The 5' ESTs were prepared from
XX total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
XX sequences usually correspond mainly to the 3' untranslated region (UTR)
XX of the mRNA because they are often obtained from oligo-dT primed cDNA
XX libraries. Such ESTs are not well suited for isolating cDNA sequences
XX derived from the 5' ends of mRNAs and even in those cases where longer
XX cDNA sequences have been obtained, the full 5' UTR is rarely included.
XX 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
XX used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
XX in diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX They are used to obtain upstream regulatory sequences and to design
XX expression and secretion vectors.
XX
XX Sequence 246 BP; 70 A; 63 C; 71 G; 40 T; 2 other;
XX
XX Query Match 4.6%; Score 131; DB 21; Length 246;
XX Best Local Similarity 100.0%; Pred. No. 1.6e-52;
XX Matches 131; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 2722 GACCCAGGCTCTCCCTGGCAAAAGATGAGACCATGAGAAACCTGGAATCCAGGCAACCA 2781
XX |||||
XX Db 112 GACCCAGGCTCTCCCTGGCAAAAGATGAGACCATGAGAAACCTGGAATCCAGGCAACCA 271
XX
XX 2782 GGGCCCCCAGGCACTTGGCAACCCATCTATGTTTATGTTATTCGAGAGAGATCCG 2841
XX |||||
XX Db 172 GGGCCCCCAGGCACTTGGCAACCCATCTATGTTTATGTTATTCGAGAGAGATCCG 231
XX
XX 2842 TTCAGAAAAGG 2852
XX |||||
XX Db 232 TTCAGAAAAGG 242
XX
XX RESULT 19
XX ABA69546/C
XX ID ABA69546 standard; DNA; 181 BP.
XX
XX AC ABA69546;
XX
XX 31-FEB-2002 (first entry)
XX
XX Human foetal liver single exon nucleic acid probe #17851.
XX
XX Human foetal liver single exon nucleic acid probe #17851.
XX
XX Human: foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
XX Homo sapiens.
XX
XX MO2001:57277-A2.
XX
XX 09-AUG-2001.

```

```

XX
XX 30-JAN-2001; 2001MO-US00669.
XX
XX 04-FEB-2000; 2000US-0180312.
XX
XX 26-MAY-2000; 2000US-C207456.
XX
XX 30-JUN-2000; 2000US-0608408.
XX
XX 03-AUG-2000; 2000US-0632366.
XX
XX 21-SEP-2000; 2000US-0234687.
XX
XX 27-SEP-2000; 2000US-0236359.
XX
XX 04-OCT-2000; 2000GB-C024263.
XX
XX (XOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Harzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-483447/52.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human fetal liver -
XX
XX Claim 4; SEQ ID NO 17851; 639bp + sequence listing; English.
XX
XX The invention relates to a single exon nucleic acid probe for
XX measuring human gene expression in a sample derived from human foetal
XX liver. The single exon nucleic acid probes may be used for predicting,
XX measuring and displaying gene expression in samples derived from human
XX fetal liver. The present sequence is a single exon nucleic acid
XX probe of the invention.
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WFO at ftp.wipo.int/pub/published_pat_sequences.
XX
XX Sequence 181 BP; 46 A; 45 C; 35 G; 55 T; 0 other;
XX
XX Query Match 2.9%; Score 82; DB 22; Length 181;
XX Best Local Similarity 100.0%; Pred. No. 6.8e-29;
XX Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 2598 AGGCTCTACGAGAAAGATGGGAAAGGAGCCAGGATTTGGGTATCTCGAGACCA 2657
XX |||||
XX Db 165 AGGCTCTACGAGAAAGATGGGAAAGGAGCCAGGATTTGGGTATCTCGAGACCA 105
XX
XX 2658 AGGCTCTCTGTCCTCCCAAGT 2679
XX |||||
XX Db 105 AGGCTCTCTGTCCTCCCAAGT 84
XX
XX RESULT 20
XX ABR71706
XX ID ABR71706 standard; cDNA; 1101 BP.
XX
XX AC ABR71706;
XX
XX 30-JUL-2002 (first entry)
XX
XX Human: dthp polynucleotide #172.
XX
XX Human; dthp; diagnostic and therapeutic polynucleotide; gene; ss; bone;
XX cell proliferative disorder; cancer; tumour; autoimmune disorder; brain;
XX inflammatory disorder; viral infection; bacterial infection; seizure;
XX fungal infection; parasitic infections; developmental disorder; breast;
XX endocrine disorder; metabolic disorder; neurological disorder; cervix;
XX gastrointestinal disorder; transport disorder; gene therapy; kidney;
XX adrenal gland; bone marrow; lung; ovary; pancreas; prostate; spleen;
XX skin; testis; thymus.
XX
XX Homo sapiens.
XX
XX MO200223754-A2.
XX
XX 14-MAR-2002.
XX
XX 29-AUG-2001; 2001MO-US27127.

```

XX PN W09845435-A2.  
 XX PD 15-OCT-1998.  
 XX PF 10-APR-1998; 98MO-US06954.  
 XX PR 10-APR-1997; 97US-0835913.  
 XX PA (GENE) GENETICS INST INC.  
 PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;  
 PI Racie LA, Spaulding V, Treacy M;  
 XX WPI; 1999-070076/06.  
 XX DR  
 XX PT New polynucleotides encoding human secreted proteins - derived from  
 PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,  
 PT ovary, pituitary, retina and colon cDNA libraries  
 XX  
 XX PS Claim 1; Page 180; 633pp; English.  
 XX CC This sequence represents an expressed sequence tag (EST), and is a  
 CC polynucleotide of the invention. The polynucleotides of the invention are  
 CC all secreted EST sequences isolated from a variety of human tissue  
 CC sources. The EST sequences and proteins encoded by them are predicted to  
 CC have useful biological activities which would make them suitable for  
 CC treating, preventing or ameliorating medical conditions in humans and  
 CC animals, although no supporting data is given. Suggested activities  
 CC include nutritional activity, immune stimulating or suppressing activity,  
 CC haematopoiesis regulating activity, tissue growth activity,  
 CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory  
 CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition  
 CC activity. The EST sequences are also stated to be useful for gene  
 CC therapy.  
 XX CC  
 XX SQ Sequence 447 BP; 141 A; 80 C; 101 G; 125 T; 0 other;  
 Query Match 8.8%; Score 252; DB 20; Length 447;  
 Best Local Similarity 99.4%; Pred. No. 7.6e-111;  
 Matches 352; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGGCTCACTATATACATTTCTCTGCATGGTTTGGTGTCTTCTTCAGAAATTCGTG 60  
 Db 85 ATGGCTCACTATATACATTTCTCTGCATGGTTTGGTGTCTTCTTCAGAAATTCGTG 144  
 QY 61 TTAGCTGAAGATGGGAAAGTAAGATCAAGTTGTGTACTGCTCCGACAGATTAGTTTC 120  
 Db 145 TTAGCTGAAGATGGGAAAGTAAGATCAAGTTGTGTACTGCTCCGACAGATTAGTTTC 204  
 QY 121 ATCTAGATGGCTTTATAGTGTGTCGCCAGAAAATTGAAATAGTGAATAAGTGCCT 180  
 Db 205 ATCTAGATGGCTTTATAGTGTGTCGCCAGAAAATTGAAATAGTGAATAAGTGCCT 264  
 QY 181 GTCAATATCAAAAAAATTGACATAGAGGCCGAAAGTTTATTCAGTTGAGTGTCAA 240  
 Db 265 GTCAATATCAAAAAAATTGACATAGAGGCCGAAAGTTTATTCAGTTGAGTGTCAA 324  
 QY 241 TATAGTACTACCTGCTGTGGAGATTCCTCTCCGAACTGTATTCAGAGGAACATTG 300  
 Db 325 TATAGTACTACCTGCTGTGGAGATTCCTCTCCGAACTGTATTCAGAGGAACATTG 384  
 QY 301 ACGGACAGAGTGAATCATACTACTTACTTAGAGGAAACACAAAAGACAGGGAG 354  
 Db 385 ACGGACAGAGTGAATCATACTACTTACTTAGAGGAAACACAAAAGACAGGGAG 418

RESULT 17  
 AA149752/C  
 ID AA149752 standard; DNA; 221 BP.  
 XX AC AA149752;

XX DT 17-OCT-2001 (first entry)  
 XX LE Probe #18438 used to measure gene expression in human placenta sample.  
 XX KW Probe; microarray; human; placenta; antenatal diagnosis;  
 XX genetic disorder; ss.  
 XX OS Home sapiens.  
 XX PK W0200157272-A2.  
 XX PD 09-AUG-2001.  
 XX PF 30-JAN-2001; 2001MO-US00663.  
 XX PR 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632386.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0226359.  
 PR 04-OCT-2000; 2000US-0024263.  
 XX PA (MOLE-) MOLECULAR DYNAMICS INC.  
 PI Bern SG, Hanzel DK, Chen W, Rank DR;  
 PI WPI; 2001-488897/53.  
 XX DR  
 XX PT Human genome-derived single exon nucleic acid probes useful for  
 PT analyzing gene expression in human placenta -  
 XX  
 XX PS Claim 25; SEQ ID NO: 18438; 654pp; English.  
 XX CC The present invention relates to single exon nucleic acid probes (SENPs).  
 CC CC The present sequence is one such probe. The probes are useful for  
 CC producing a microarray for predicting, measuring and displaying gene  
 CC expression in samples derived from human placenta. The probes are useful  
 CC for antenatal diagnosis of human genetic disorders.  
 XX CC  
 XX SQ Sequence 221 BP; 62 A; 59 C; 43 G; 57 T; 0 other;  
 Query Match 5.8%; Score 166; DB 22; Length 221;  
 Best Local Similarity 100.0%; Pred. No. 2.2e-69;  
 Matches 166; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2435 GAATTGTGATCATTGCTCTGTCCCAACAATGCTCCCGGATATTCCTGSSCCACCTGGTC 2494  
 Db 22: GAATTGTGATCATTGCTCTGTCCCAACAATGCTCCCGGATATTCCTGSSCCACCTGGTC 162  
 QY 2495 CGATAGGCCCAAGAGGCTCCGAGAGATTACCTGTTGCGAGAGAGATGGTTCCTG 2554  
 Db 161 CGATAGGCCCAAGAGGCTCCGAGAGATTACCTGTTGCGAGAGAGATGGTTCCTG 102  
 QY 2555 GATTAGTGGGTGCTCCTGAGCTCCAGCTGCAGAGGATTAAAGC 2600  
 Db 101 GATTAGTGGGTGCTCCTGAGCTCCAGCTGCAGAGGATTAAAGG 56

RESULT 18  
 AAC29851  
 ID AAC29851 standard; cDNA; 246 BP.  
 XX AC AAC29851;  
 XX DT 06-OCT-2000 (first entry)  
 XX DE Human secreted protein; 5' EST; SEQ ID NO: 33926.  
 XX KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
 KW gene therapy; chromosome mapping; ss.  
 XX

QY 781 TCAAAAGTATATACAGAACTCACAAGCAATGTTTCCACAGAGGCTTCTCCATCA 840  
 |||||  
 Db 975 TCAAAAGTATATATACAGAACTCACAAGCAATGTTTCCACAGAGGCTTCTCCATCA 1034  
 |||||  
 QY 841 TATGATTTGTGCTACTCAAAAGATTTAAAGTCAGAAAATTTGGATTATGAGAAATA 900  
 |||||  
 Db 1035 TATGATTTGTGCTACTCAAAAGATTTAAAGTCAGAAAATTTGGATTATGAGAAATA 1094  
 |||||  
 QY 901 TTAATCTATGATGGAAGGCCA 921  
 |||||  
 Db 1095 TTAATCTATGATGGAAGGCCA 1115  
 |||||  
 RESULT 15  
 ABR71703  
 ID ABR71703 standard; cDNA; 1414 BP.  
 XX  
 AC ABR71703;  
 XX  
 DT 30-JUL-2002 (first entry)  
 XX  
 CE Human dithp polynucleotide #169.  
 XX  
 KW Human; dithp; diagnostic and therapeutic polynucleotide; gene; ss; bone;  
 KW cell proliferative disorder; cancer; tumour; autoimmune disorder; brain;  
 KW inflammatory disorder; viral infection; bacterial infection; seizure;  
 KW fungal infection; parasitic infections; developmental disorder; breast;  
 KW endocrine disorder; metabolic disorder; neurological disorder; cervix;  
 KW gastrointestinal disorder; transport disorder; gene therapy; kidney;  
 KW adrenal gland; bone marrow; lung; ovary; pancreas; prostate; spleen;  
 KW skin; testis; thymus.  
 XX  
 OS Homo sapiens.  
 XX  
 CN MO200220754 A2.  
 PD 14-MAR-2002.  
 XX  
 PF 29-AUG-2001; 2001MO-US27127.  
 XX  
 PR 05-SEP-2000; 2000US-229747P.  
 PR 05-SEP-2000; 2000US-229748P.  
 PR 05-SEP-2000; 2000US-229749P.  
 PR 05-SEP-2000; 2000US-229750P.  
 PR 05-SEP-2000; 2000US-229751P.  
 PR 05-SEP-2000; 2000US-230583P.  
 PR 06-SEP-2000; 2000US-230585P.  
 PR 06-SEP-2000; 2000US-230514P.  
 PR 06-SEP-2000; 2000US-230515P.  
 PR 06-SEP-2000; 2000US-230517P.  
 PR 06-SEP-2000; 2000US-230518P.  
 PR 06-SEP-2000; 2000US-230519P.  
 PR 06-SEP-2000; 2000US-230595P.  
 PR 06-SEP-2000; 2000US-230597P.  
 PR 06-SEP-2000; 2000US-230598P.  
 PR 06-SEP-2000; 2000US-230599P.  
 PR 06-SEP-2000; 2000US-230610P.  
 PR 06-SEP-2000; 2000US-230865P.  
 PR 06-SEP-2000; 2000US-230988P.  
 PR 07-SEP-2000; 2000US-230951P.  
 PR 07-SEP-2000; 2000US-231167P.  
 PR 07-SEP-2000; 2000US-231167P.  
 XX  
 PA (INCYTE GENOMICS INC.  
 XX  
 PI Stuart J, Lincoln SE, Altus CM, Dufour GE, Chalup KS, Hillman JT,  
 PI Jones AL, Yu JY, Wright RJ, Gierzen D, Liu TF, Yap PS, Dahl CR,  
 PI Momiyama MG, Bradley DL, Rohatgi SD, Harris B, Roseberry AM,  
 PI Gerstein RH, Peralta CH, David WF, Panzer SR, Flores V, Daffo A,  
 PI Kawata R, Chen AJ, Chang SC, Au AP, Inman RR;  
 XX  
 DR WPI; 2002-383054/41.  
 DR P-PSDB; ABG60112.

XX  
 PT An isolated polynucleotide useful in diagnostics and therapeutics -  
 XX  
 PS Claim 1: Page 486; 686pp; English.  
 XX  
 CC The invention relates to human diagnostic and therapeutic (dithp)  
 CC polynucleotides and their associated polypeptides (dithp polypeptides).  
 CC The sequences of the invention are used in the treatment and diagnosis of  
 CC cell proliferative disorders (e.g. atherosclerosis, cirrhosis), cancers  
 CC (e.g. tumours of the adrenal gland, bone, bone marrow, brain, breast,  
 CC cervix, kidney, lung, ovary, pancreas, prostate, skin, spleen, testis or  
 CC thymus), autoimmune/inflammatory disorders (e.g. asthma, bronchitis,  
 CC psoriasis, osteoporosis), viral infections, bacterial infections, fungal  
 CC infections, parasitic infections, developmental disorders (e.g. anaemia,  
 CC epilepsy), seizure disorders (e.g. cerebral palsy, spina bilida),  
 CC endocrine disorders (e.g. thrombosis, aneurysm), metabolic disorders  
 CC (e.g. obesity, diabetes), neurological disorders (e.g. stroke,  
 CC amyotrophic lateral sclerosis, multiple sclerosis), gastrointestinal  
 CC disorders (e.g. ulcerative colitis, lymphoma) and transport disorders  
 CC (e.g. myotonic dystrophy, catatonid, peripheral neuropathy). Sequences  
 CC ABR71535-ABR71809 represent human dithp polynucleotides of the invention.  
 XX  
 SC Sequence 1414 BP; 425 A; 267 C; 295 G; 427 T; 0 other;  
 Query Match 13.3%; Score 382; DB 24; Length 1414;  
 Best Local Similarity :00.0%; Pred. No. 1.7e-173;  
 Matches 382; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2397 AGCCGAGTACCACTCTTACTTCGAGTGAAGAATTGAATTGATGATTCCTGTC 2456  
 |||||  
 Db 959 AGCCGAGTACCACTCTTACTTCGAGTGAAGAATTGAATTGATGATTCCTGTC 1018  
 |||||  
 QY 2457 CCACATGCTCCCGGGTATCTCTGGGCCACCGTGGCCATAGGCCCAAGGCTCCGAG 2516  
 |||||  
 Db 1019 CCACATGCTCCCGGGTATCTCTGGGCCACCGTGGCCATAGGCCCAAGGCTCCGAG 1078  
 |||||  
 QY 2517 AGGATTAAGTGTGCTGAGAGAGATGATGCTCTGATTAAGTGGGTGCTCCGAGAGC 2576  
 |||||  
 Db 1079 AGGATTAAGTGTGCTGAGAGAGATGATGCTCTGATTAAGTGGGTGCTCCGAGAGC 1138  
 |||||  
 QY 2577 TCCAGGTGTCAGAGATTTAAAGGCTTACCCAGAGAAATGGGAAAAAGGAGCCCAAG 2636  
 |||||  
 Db 1139 TCCAGGTGTCAGAGATTTAAAGGCTTACCCAGAGAAATGGGAAAAAGGAGCCCAAG 1198  
 |||||  
 QY 2637 GTTGGGATCTCTGAGAAAGGCTCTCTGGGCCCGCCAGGTCACAGAGGCCCTCTGG 2696  
 |||||  
 Db 1199 GTTGGGATCTCTGAGAAAGGCTCTCTGGGCCCGCCAGGTCACAGAGGCCCTCTGG 1258  
 |||||  
 QY 2697 AATAAGCAAGAGGCTCTCCAGAGAGCCAGGTCCTCCAGCAAGATGAGAGCATGG 2756  
 |||||  
 Db 1259 AATAAGCAAGAGGCTCTCCAGAGAGCCAGGTCCTCCAGCAAGATGAGAGCATGG 1318  
 |||||  
 QY 2757 AAAACCTGGAATCCAAAGGCCA 2778  
 |||||  
 Db 1319 AAAACCTGGAATCCAAAGGCCA 1340  
 |||||  
 RESULT 16  
 AA986274  
 ID AA986274 standard; cDNA; 447 BP.  
 XX  
 AC AA986274;  
 XX  
 DT 27-APR-1999 (first entry)  
 XX  
 DE EST clone AJ54.  
 XX  
 KW Expressed sequence tag; secreted protein; haematopoiesis regulator;  
 KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;  
 KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolytic;  
 KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.  
 XX  
 OS Homo sapiens.

KM acquired immune deficiency syndrome; AIDS; autoimmune disorder;  
 KM respiratory disorder; ss.  
 XX Homo sapiens.  
 OS Homo sapiens.  
 PN W0200162927-A2.  
 XX 30-AUG-2001.  
 XX 21-FEB-2001; 2001WO-US06059.  
 XX 24-FEB-2000; 2000US-0184697.  
 XX 24-FEB-2000; 2000US-0184697.  
 XX 24-FEB-2000; 2000US-0184698.  
 XX 24-FEB-2000; 2000US-0184768.  
 XX 24-FEB-2000; 2000US-0184769.  
 XX 24-FEB-2000; 2000US-0184770.  
 XX 24-FEB-2000; 2000US-0184771.  
 XX 24-FEB-2000; 2000US-0184772.  
 XX 24-FEB-2000; 2000US-0184773.  
 XX 24-FEB-2000; 2000US-0184774.  
 XX 24-FEB-2000; 2000US-0184776.  
 XX 24-FEB-2000; 2000US-0184777.  
 XX 24-FEB-2000; 2000US-0184797.  
 XX 24-FEB-2000; 2000US-0184813.  
 XX 24-FEB-2000; 2000US-0184813.  
 XX 24-FEB-2000; 2000US-0184837.  
 XX 24-FEB-2000; 2000US-0184841.  
 XX 24-FEB-2000; 2000US-0185213.  
 XX 24-FEB-2000; 2000US-0185216.  
 XX 12-MAY-2000; 2000US-0203785.  
 XX 15-MAY-2000; 2000US-0204226.  
 XX 16-MAY-2000; 2000US-0204525.  
 XX 16-MAY-2000; 2000US-0204821.  
 XX 16-MAY-2000; 2000US-0204908.  
 XX 16-MAY-2000; 2000US-0205232.  
 XX 17-MAY-2000; 2000US-0204813.  
 XX 17-MAY-2000; 2000US-0204813.  
 XX 17-MAY-2000; 2000US-0205221.  
 XX 17-MAY-2000; 2000US-0205285.  
 XX 17-MAY-2000; 2000US-0205286.  
 XX 17-MAY-2000; 2000US-0205287.  
 XX 17-MAY-2000; 2000US-0205323.  
 XX 17-MAY-2000; 2000US-0205324.  
 XX (INCY-) INCYTE GENOMICS INC.  
 XX Panzer SR, Spiro PA, Banville SC, Shah P, Chalup MS, Chang SC;  
 XX Chen A, D'Sa SA, Amshay S, Dahl CR, Dam TC, Daniels SE;  
 XX Dufour GE, Flores V, Fong WT, Greenwalt JB, Hillman JL, Jones AL;  
 XX Liu TF, Roseberry AM, Rosen BH, Russo FD, Stockdresher TK, Datto A;  
 XX Wright RJ, Yap PE, Yu JY, Bradley DJ, Bratcher SR, Chert W;  
 XX Cohen HJ, Hodgson DM, Lincoln SE, Jackson S;  
 XX WPI: 2001-502867/55.  
 XX P-PSDB: AAU19561.  
 XX Polynucleotides encoding diagnostic and therapeutic proteins, e.g.  
 XX enzymes, hormones and receptors, useful in diagnostics and therapeutics  
 XX -  
 XX Claim 1; Page 365; 522pp; English.

CC leukaemia, autoimmune disorders, and respiratory disorders. Additionally,  
 CC (I) may be used to produce the DTHPs, by inserting the nucleic acids  
 CC into a host cell and culturing the cell to express the protein. (I) and  
 CC its complementary sequences may also be used as DNA probes in diagnostic  
 CC assays to detect and quantitate the presence of similar nucleic acids in  
 CC samples, and therefore which patients may be in need of restorative  
 CC therapy. (II) may also be used as antigens in the production of DTHP  
 CC antibodies against DTHPs and in assays to identify modulators of DTHP  
 CC expression and activity. The anti-DTHP antibodies and antagonists may  
 CC also be used to down regulate expression and activity. The anti-DTHP  
 CC antibodies may also be used as diagnostic agents for detecting the  
 CC presence of DTHPs in samples (e.g. by enzyme linked immunosorbent  
 CC assay (ELISA)). AAS30986-AAS31196 represent human diagnostic and  
 CC therapeutic (DTHP) polynucleotides of the invention.  
 CC  
 XX Sequence 1115 BP; 344 A; 215 C; 259 G; 296 T; 1 other;  
 SQ  
 Query Match 26.1%; Score 749; DB 22; Length 1115;  
 Best Local Similarity 99.8%; Pred. No. 0;  
 Matches 919; Conservative 0; Mismatches 1; Indels 1; Gaps 1;  
 QY 1 ATGGCTCATATATACATTTCTGCAATGTTGGTGGCTTCTTCAAAATTCGTG 60  
 ob ATGGCTCATATATACATTTCTGCAATGTTGGTGGCTTCTTCAAAATTCGTG 255  
 QY 61 TTAGCTGAAGTGGGAGAGTCAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 120  
 Db TTAGCTGAAGTGGGAGAGTCAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 315  
 QY 121 ATCTTAGATGGCTCTTATAGTGTGGCCGCAAAACTTGAATAGTGAATAGTGGCTT 180  
 Db ATCTTAGATGGCTCTTATAGTGTGGCCGCAAAACTTGAATAGTGAATAGTGGCTT 375  
 QY 181 GTCAATATCAAAAAAATTTGACATAGGCGGCAAGTTATTCANATTTGAGTGGTCA 240  
 Db GTCAATATCAAAAAAATTTGACATAGGCGGCAAGTTATTCANATTTGAGTGGTCA 435  
 QY 241 TATAGTACCTACCCCTGGTGGAGATCTCTCGGAAAGTATGATTCAGGAGAACTTTG 300  
 Db TATAGTACCTACCCCTGGTGGAGATCTCTCGGAAAGTATGATTCAGGAGAACTTTG 495  
 QY 301 ACAGCAGCATGTGCAATTCATCTACTTAGAGAGAAACACAAAGCAGGAGGCGCATC 360  
 Db ACAGCAGCATGTGCAATTCATCTACTTAGAGAGAAACACAAAGCAGGAGGCGCATC 555  
 QY 361 CAGTTGGCTGGATTAATCTTTTGGCAAGTCCACAGATTCCTGACATAGATAGCATG 420  
 Db CAGTTGGCTGGATTAATCTTTTGGCAAGTCCACAGATTCCTGACATAGATAGCATG 615  
 QY 421 GTACTTAGAGATGGCAATTCACAGATGACGTCAGAGATGCAAGTCAAGCAGCAAGAT 480  
 Db GTACTTAGAGATGGCAATTCACAGATGACGTCAGAGATGCAAGTCAAGCAGCAAGAT 675  
 QY 481 AGTAGAGATTAATTTTGTCTATGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 540  
 Db AGTAGAGATTAATTTTGTCTATGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 735  
 QY 541 GGTATGGCAACAGCCCTGCTACTATGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 600  
 Db GGTATGGCAACAGCCCTGCTACTATGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 795  
 QY 601 TCCAAATTAAGGGAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660  
 Db TCCAAATTAAGGGAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 855  
 QY 661 ATTCAGTGGCAGCTCGGATGAAGGGAATGATATTTCTTATAGTGGTGGTGGTGGTGGT 720  
 Db ATTCAGTGGCAGCTCGGATGAAGGGAATGATATTTCTTATAGTGGTGGTGGTGGTGGT 914  
 QY 721 AAAAGGTTAAGAAAGATACAGCTTTCACCAAAAAAATAAAGAGATATGAAGTACA 780  
 Db AAAAGGTTAAGAAAGATACAGCTTTCACCAAAAAAATAAAGAGATATGAAGTACA 974

PR 06-SEP-2000; 2000US-230599P.  
 PR 06-SEP-2000; 2000US-230610P.  
 PR 06-SEP-2000; 2000US-230865P.  
 PR 06-SEP-2000; 2000US-230988P.  
 PR 06-SEP-2000; 2000US-230989P.  
 PR 07-SEP-2000; 2000US-230951P.  
 PR 07-SEP-2000; 2000US-231163P.  
 PR 07-SEP-2000; 2000US-231167P.  
 XX (INCYT) INCYTE GENOMICS INC.

PA Jackson S, Lincoln SE, Altus CM, Dufour GE, Chalup MS, Hillman JL;  
 PI Jones AL, Yu JY, Wright RJ, Gierzen D, Liu TF, Yap PE, Dahl CR;  
 PI Montiyama MG, Bradley DL, Rohatgi SD, Harris B, Roseberry AK;  
 PI Gerstin EH, Peralta CH, David MH, Panzer SR, Flores V, Daffo A;  
 PI Marwaha R, Chen AU, Chang SC, Au AP, Imman RR;  
 XX MPI: 2002-527544/56.  
 DR P-PSDB; ABP51293.

PT Novel human disease detection and treatment polypeptide, useful in  
 PT diagnosis, prevention or treatment of cell proliferative disorders e.g.  
 PT arteriosclerosis, cirrhosis and an autoimmune/inflammatory disorder  
 PT e.g. AIDS

XX Claim 1; Page 341; 618pp; English.

XX The invention relates to an isolated human disease detection and  
 CC treatment (MDT) polypeptide (I) selected from a polypeptide having a  
 CC sequence selected from 254 sequences (ABP51231-ABP51484) given in the  
 CC specification, a naturally occurring polypeptide comprising a sequence  
 CC having at least 90% identity to (I) or a biologically active or  
 CC immunogenic fragment of (I). (I) is useful for screening a compound for  
 CC effectiveness as an agonist or antagonist for screening a compound that  
 CC specifically binds (II) or modulates the activity of (I), and for  
 CC preparing a polyclonal or monoclonal antibody by hybridoma technology.  
 CC Nucleic acids (II) (ABQ72449-ABQ72700) encoding (I) are useful for  
 CC screening a compound for effectiveness in altering expression of a target  
 CC polynucleotide comprising. Oligonucleotides and antibodies are useful for  
 CC detecting MDDT in a sample or for assessing toxicity of a test compound,  
 CC in a diagnostic test for a condition or a disease associated with the  
 CC expression of MDDT in a biological sample, for detecting (I) in a sample,  
 CC and for purifying (I) from a sample. A composition comprising (I), an  
 CC agonist or antagonist is useful for treating a disease or condition  
 CC associated with decreased or increased expression of functional MDDT.  
 CC (I) or (II) are useful for diagnosing, treating or preventing disorders  
 CC associated with aberrant expression of MDDT, where the disorders are  
 CC selected from a cell proliferative disorder such as arteriosclerosis,  
 CC cirrhosis, hepatitis, psoriasis, and cancer and an  
 CC autoimmune/inflammatory disorder such as AIDS, Addison's disease,  
 CC allergy, anaemia, asthma, atherosclerosis, gout, multiple sclerosis or  
 CC rheumatoid arthritis. (II) are useful for creating knockin humanised  
 CC animals or transgenic animals to model human diseases, in somatic or  
 CC germ-line gene therapy, to generate a transcript image of a tissue or cell  
 CC type, for detecting differences in the chromosomal location due to  
 CC translocation or inversion among normal, carrier or affected individuals  
 CC and as hybridisation probes for mapping naturally occurring genomic  
 CC sequences.

XX Sequence 1169 BP; 364 A; 224 C; 267 G; 314 T; 0 other;

XX Query Match 32.2%; Score 923; DB 24; Length 1169;  
 XX Best Local Similarity 99.9%; Pred. No. 0;  
 XX Matches 973; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGGCTCACTAATTACATTTCTGCGATGGTTTGGTGGCTCTTCAGAAATCTGTG 60  
 DB 196 ATGGCTCACTAATTACATTTCTGCGATGGTTTGGTGGCTCTTCAGAAATCTGTG 255  
 QY 61 TTAGCTTAAGATGGGAGAGTAAGATCAAGTTGCTGACTGCTCCGACATTTAGTTTTC 120  
 DB 256 TTAGCTTAAGATGGGAGAGTAAGATCAAGTTGCTGACTGCTCCGACATTTAGTTTTC 315

QY 121 ATCTTAGATGCGCTCTTATATAGTGTGGCCCGAGAAAACCTTGAATATGTGAAAAAGTGGCTT 180  
 DB 316 ATCTTAGATGCGCTCTTATAGTGTGGCCCGAGAAAACCTTGAATATGTGAAAAAGTGGCTT 375  
 QY 181 GTCAATATGCAAAAAAAGTTTACATATAGGCGGAAAGTTATATCAAGTTGAGAGTGGTCAA 240  
 DB 370 GTCAATATGCAAAAAAAGTTTACATATAGGCGGAAAGTTATATCAAGTTGAGAGTGGTCAA 435  
 QY 241 TATAGTACATACCTCTGCTGAGATTCCTCGAAGCTATGATTCAGAGAACATTTTG 300  
 DB 436 TATAGTACATACCTCTGCTGAGATTCCTCGAAGCTATGATTCAGAGAACATTTTG 495  
 QY 301 ACCGACAGATGGAATCCATATCTTATAGAGAAAACAAAGACAGAGAGGCGATC 360  
 DB 496 ACCGACAGATGGAATCCATATCTTATAGAGAAAACAAAGACAGAGAGGCGATC 555  
 QY 361 CAGTTTGGCTGATACCTTTTTCGCAAGTCCCTGCAATTTCTGACTAATAGACAGTG 420  
 DB 556 CAGTTTGGCTGATACCTTTTTCGCAAGTCCCTGCAATTTCTGACTAATAGACAGTG 615  
 QY 421 GTACTTAGGATGGCAATCCAGATGACGTCAAGATGACAGTCCAGCAGCAAGAT 480  
 DB 616 GTACTTAGGATGGCAATCCAGATGACGTCAAGATGACAGTCCAGCAGCAAGAT 675  
 QY 481 AGTACGACACATGATTTGGTATGGTGGTTCGAAACAGAAAGATGCGAATCTTAA 540  
 DB 676 AGTACGACACATGATTTGGTATGGTGGTTCGAAACAGAAAGATGCGAATCTTAA 735  
 QY 541 GCTATTGCGCAAGAGCCTTCGTCTACTTATGTTTATATGTAAGACATATATTCGCAATA 600  
 DB 736 GCTATTGCGCAAGAGCCTTCGTCTACTTATGTTTATATGTAAGACATATATTCGCAATA 795  
 QY 601 TCCAAATTAAGGAGAGTGAAGAGCAAGAACTTTGAGAGATCTGCTGCCACACAGA 660  
 DB 796 TCCAAATTAAGGAGAGTGAAGAGCAAGAACTTTGAGAGATCTGCTGCCACACAGA 855  
 QY 661 ATTCAGATGGCAGCTGATGATGAAGAGGATTTGATATCTTTAGTGTAGATGTAAAT 720  
 DB 856 ATTCAGATGGCAGCTGATGATGAAGAGGATTTGATATCTTTAGTGTAGATGTAAAT 915  
 QY 721 AAAAAGTTAAGAAAAGATATACGCTTTTACCCAAAATAAAGATATGAAGTACTA 780  
 DB 916 AAAAAGTTAAGAAAAGATATACGCTTTTACCCAAAATAAAGATATGAAGTACTA 975  
 QY 780 TCAAAAGTTATTTATACAGAACTCAACAGCAATGTTTCCGAAAGTGTTCCTCATCA 840  
 DB 976 TCAAAAGTTATTTATACAGAACTCAACAGCAATGTTTCCGAAAGTGTTCCTCATCA 1035  
 QY 841 TATGATTTTGTGCTTACTCAAGATTTAAAGTCAAGAAAATTTGGGATTTATGAGAAATA 900  
 DB 1036 TATGATTTTGTGCTTACTCAAGATTTAAAGTCAAGAAAATTTGGGATTTATGAGAAATA 1095  
 QY 901 TTACTATTTATGAGAGGCGCAAAATGAGTACTCTTAATGAGTGGTGCACAAATCTTA 960  
 DB 1096 TTACTATTTATGAGAGGCGCAAAATGAGTACTCTTAATGAGTGGTGCACAAATCTTA 1155  
 QY 961 TTATTTACAAACAAC 974  
 DB 1156 TTATTTACAAACAAC 1169

RESULT 14  
 AAS31132  
 ID AAS31132 standard; CEWA; 1115 BP.  
 XX AAS31132;  
 XX 04-DEC-2001 (first entry)  
 XX Human diagnostic and therapeutic polynucleotide (DITHP) #147.  
 XX Human; receptor; diagnostic; therapeutic; gene therapy; vaccine;  
 KW cell proliferative disorder; Crohn's disease; lymphoma; leukemia;



PR 03-AUG-2000; 2000US-0653450.  
PR 14-SEP-2000; 2000US-0662191.  
PR 19-OCT-2000; 2000US-0693036.  
PR 29-NOV-2000; 2000US-0727344.

PA (HYSE-) HYSEQ INC.

P1 Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D,  
P1 Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
P1 Zhao QA, Zhou F, Goodrich R, Drmanac RT;

DR WPI; 2001-442253/47.  
DR P-PSDB; AAM41032.

Novel nucleic acids and polypeptides, useful for treating disorders such as central nervous system injuries -

PS Claim 1; SEQ ID NO 4177; 10078bp; English

The invention relates to human nucleic acids (AA157799-AA161369) and the encoded polypeptides (AA138642-AA142213) with neurotrophic, immunosuppressant and cyostatic activity. The polynucleotides are useful in gene therapy, a composition containing a polypeptide or polynucleotide of the invention may be used to treat diseases of the peripheral nervous system, such as peripheral nerves injuries, peripheral neuropathy and localised neuropathies and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager Syndrome. Other uses include the utilisation of the activities such as: Immune system suppression, Activin/inhibin activity, chemocarcin/chemokine activity, haemostatic and thrombolytic activity, cancer diagnosis and therapy, drug screening assays for receptor activity, arthritis and inflammation, leukemias and C.N.S disorders.

Note: The sequence data for this patent did not form part of the printed specification.

52 Sequence 2209 BP; 674 A; 413 C; 544 G; 577 T; 1 other;

Query Match 33.7%; Score 966; DB 22; Length 2209;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 966; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY	1900	CCCTGGTTTAAATGGGAAGCAATGCTCTACCAAGGCCACGCTGGAAACCCCGGAGTTTAAAGGA	195
Pb	167	CCCTGGTTTAAATGGGAAGCAATGGCTCAACAGGCCACGCTGGAAACCCCGGAGTTTAAAGGA	226
QY	1960	AGCAAAAGTGAACTTGAAATTCAGAGGATGGCTCGGGCTTCTGGGCTCAAGAGAGAACCA	2218
Pb	227	AGCAAAAGTGAACTTGAAATTCAGAGGATGGCTCGGGCTTCTGGGCTCAAGAGAGAACCA	286
QY	2020	GGAGCAACGGGTTCCCCAGAGAACCAAGATACATGGGTTTACCCGGATTCAGAGAAA	2079
Pb	287	GGAGCAACGGGTTCCCCAGAGAACCAAGATACATGGGTTTACCCGGATTCAGAGAAA	346
QY	2080	AAAGGGGACAAAGGAAATTCAGGTCGTAAGGTAATCTAGCGTCAAAAGGAGAAATGCA	2133
Pb	347	AAAGGGGACAAAGGAAATTCAGGTCGTAAGGTAATCTAGCGTCAAAAGGAGAAATGCA	406
QY	2140	AGA CAGGGAATTCAGGCGCAACAGGGAATTCAGGCGCATATGTCGTAAGAGAGAGA	2193
Pb	407	AGACACAGGAAATTCAGGCGCAACAGGGAATTCAGGCGCATATGTCGTAAGAGAGAGA	466
QY	2200	GGTGAAGAGGAAACCTGGTCTCCGAGGTCGCAATTGGAATCAAAAGAGAAATTCGGGGTG	2255
Pb	467	GGTGAAGAGGAAACCTGGTCTCCGAGGTCGCAATTGGAATCAAAAGAGAAATTCGGGGTG	526
QY	2260	GATGCTTATGGGGCCGACAGTCTTAAGGGGCAACCTGGGAGTCAGGTCCTCAGGGA	2318
Pb	527	GATGCTTATGGGGCCGACAGTCTTAAGGGGCAACCTGGGAGTCAGGTCCTCAGGGA	586
QY	2320	CCCCCAGGTTTGATGGAGACCCCGGAAGAGGATTTTCAGACAAATTTATTCACAGATT	2375
Pb	587	CCCCCAGGTTTGATGGAGACCCCGGAAGAGGATTTTCAGACAAATTTATTCACAGATT	646

QY	2380	TGCACAGATGTAATTAAGGCCCGACGTACCAAGCTTATCTTCAGAGTGGAAAGAAATTAAGAAAT	2439
Db	647	TCCACAGATGTAATTAAGGCCCGACGTACCAAGCTTATCTTCAGAGTGGAAAGAAATTAAGAAAT	705
QY	2440	TGTGATCATTCCTCCTCCTCCCAACATAGCTCCCGGGATATTCCTGGCCCACTGGTCCGA7A	2499
Db	707	TGTGATCATTCCTCCTCCTCCCAACATAGCTCCCGGGATATTCCTGGCCCACTGGTCCGA7A	766
QY	2530	GGCCCCAGAGGGTCCCAAGAGATTAACCTGGTTTCCACAGAGAAAGATAGTGTCTCCCGATTA	2559
Db	767	GGCCCCAGAGGGTCCCAAGAGATTAACCTGGTTTCCACAGAGAAAGATAGTGTCTCCCGATTA	826
QY	2560	GTGGGTGTCCCTGACAGCTCCAGAGTGTACAGAGATTTAAAGAGCCCTTCCAGSAAAGAAATGGG	2619
Db	827	GTGGGTGTCCCTGACAGCTCCAGAGTGTACAGAGATTTAAAGAGCCCTTCCAGSAAAGAAATGGG	886
QY	2620	GAAGAAAGGAGCCCAAGGGTTTGGGTATCTCTGAGAAACAGAGTCTCTCTGGTCCCCAGCT	2679
Db	887	GAAGAAAGGAGCCCAAGGGTTTGGGTATCTCTGAGAAACAGAGTCTCTCTGGTCCCCAGCT	946
QY	2680	CCAGAGAGGCCCTCTCTGATTAAGCAAGAGTCTCTCCAGAGACCAAGTCTCCTCTGAC	2739
Db	947	CCAGAGAGGCCCTCTCTGATTAAGCAAGAGTCTCTCCAGAGACCAAGTCTCCTCTGAC	1006
QY	2740	AAAGATGGAGACCATGGAAACCTGGAAATCCAGAGGCAACCAAGGCCCTCCAGGATCTGC	2799
Db	1007	AAAGATGGAGACCATGGAAACCTGGAAATCCAGAGGCAACCAAGGCCCTCCAGGATCTGC	1066
QY	2800	GACCCATCACTATGTTTATGTATATTGGCCAGAAAGATCCGTTCCAGAAAGACCAAAAC	2859
Db	1067	GACCCATCACTATGTTTATGTATATTGGCCAGAAAGATCCGTTCCAGAAAGACCAAAAC	1126
QY	2860	TATTAG 2865	
Db	1127	TATTAG 1132	
RESULT 13			
ID	ABQ72511	standard: cDNA, 1169 bp.	
XX	ABQ72511:		
AC	03-SEP-2002 (first entry)		
XX	Human MDT encoding cDNA SEQ ID NO 63.		
DE	Human, MDT, disease detection and treatment molecule polynucleotide;		
KW	proliferative disorder; hepatitis; psoriasis; cancer; AIDS;		
KW	autoimmune disorder; inflammatory disorder; allergy; multiple sclerosis;		
KW	rheumatoid arthritis; transgenic gene therapy; antiarteriosclerotic;		
KW	hepatotropic; antiinflammatory; antiparasitic; cytostatic; anti-HIV;		
KM	antiallergic; antineurotic; antineurotic; antineurotic; antineurotic;		
XX	neuroprotective; antineurotic; antineurotic; antineurotic; antineurotic;		
CS	homo sapiens.		
XX	W02002040715-A2.		
PN	23-MAY-2002.		
PD	06-SEP-2001; 2001-WO-US27628.		
PF	06-SEP-2000; 2000US-230505P.		
PR	06-SEP-2000; 2000US-230514P.		
PR	06-SEP-2000; 2000US-230515P.		
PR	06-SEP-2000; 2000US-230517P.		
PR	06-SEP-2000; 2000US-230518P.		
PR	06-SEP-2000; 2000US-230519P.		
PR	06-SEP-2000; 2000US-230597P.		
PR	06-SEP-2000; 2000US-230598P.		

PD 09-AUG-2001.  
 XX 05-FEB-2001; 2001WO-US04098.  
 XX 03-FEB-2000; 2000US-0496914.  
 PR 27-APR-2000; 2000US-0550875.  
 PR 20-JUN-2000; 2000US-0598075.  
 PR 19-JUL-2000; 2000US-0620325.  
 PR 01-SEP-2000; 2000US-0654936.  
 PR 15-SEP-2000; 2000US-0663561.  
 PR 20-OCT-2000; 2000US-0693125.  
 PR 30-NOV-2000; 2000US-0728422.  
 XX (HSE-) HSEQ INC.  
 PA Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;  
 PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZM;  
 PI Xue AJ, Yang Y, Wehrman T, Goodrich R;  
 XX WPI: 200:476283/51.  
 DR P-PS08; AAM79896.  
 DR Nucleic acids encoding polypeptides with cytokine-like activities,  
 PT useful in diagnosis and gene therapy -  
 XX Claim 1; Page 4838; 6221pp; English.  
 CC The invention relates to polynucleotides (AAK51456-AAK51435) and the  
 CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity relating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoietic regulatory  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activin/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukemia, nervous system disorders, arthritis and  
 CC inflammation.  
 CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666  
 CC (AAM80020) are omitted as the relevant pages from the sequence listing  
 CC were missing at the time of publication.  
 CC  
 XX  
 XX Sequence 2209 BP; 674 A; 413 C; 544 G; 577 T; 1 other;  
 SO  
 Query Match 33.7%; Score 966; DB 22; Length 2209;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 966; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 527 GATGGCTTGATGGGGCCCCAGGCTCTAAGGGGCAACCTGSGAGATCGAGTCTCAGGGA 586  
 QY 2320 CCCCAGGTTGGATGGGAAGCCCGAAGAGTTTCAAGCAATTTATTCAGAGTT 2379  
 DB 587 CCCCAGGTTGGATGGGAAGCCCGAAGAGTTTCAAGCAATTTATTCAGAGTT 646  
 QY 2380 TGCACAGATTAATTAAGAGCCGAGTACAGTCTTAATTCAAGAGTGAAGAAATTGAAT 2439  
 DB 647 TGCACAGATTAATTAAGAGCCGAGTACAGTCTTAATTCAAGAGTGAAGAAATTGAAT 706  
 QY 2440 TGTGATATTCCTGTCCTCCCAAGCATGGCTCCCCGGGATTCCTGGGCTCACTGGTCCGATA 2495  
 DB 707 TGTGATATTCCTGTCCTCCCAAGCATGGCTCCCCGGGATTCCTGGGCTCACTGGTCCGATA 766  
 QY 2500 GGCCACAGAGGTTCCAGAGATTAAGTGGTTGACAGAGAGATGGTCTTCCTGGATTA 2559  
 DB 767 GGCCACAGAGGTTCCAGAGATTAAGTGGTTGACAGAGAGATGGTCTTCCTGGATTA 826  
 QY 2560 GTGGGTGTCCTGAGAGTCCAGGTTGAGAGATTAAGAGCCCTACAGAGAAATGGG 2619  
 DB 827 GTGGGTGTCCTGAGAGTCCAGGTTGAGAGATTAAGAGCCCTACAGAGAAATGGG 886  
 QY 2620 GAAAAAGAGAGGTTGGGTTGAGAGATTAAGAGCCCTACAGAGAAATGGG 2679  
 DB 887 GAAAAAGAGAGGTTGGGTTGAGAGATTAAGAGCCCTACAGAGAAATGGG 946  
 QY 2680 CAGAGAGGCTCTCTGAGATTAAGCAAGAGTCTCCAGAGACCCAGGTTCTCTGGC 2739  
 DB 947 CAGAGAGGCTCTCTGAGATTAAGCAAGAGTCTCCAGAGACCCAGGTTCTCTGGC 1006  
 QY 2740 AAGATGAGACCATGAGAAACCTGGAATCCAAAGGCAACAGGCCCCAGGATCTGC 2799  
 DB 1009 AAGATGAGACCATGAGAAACCTGGAATCCAAAGGCAACAGGCCCCAGGATCTGC 1066  
 QY 2800 GACCATCATATGTTTATGATTAATTCAGAGAGATCCGTTGAGAAAGAGACCAAC 2859  
 DB 1067 GACCATCATATGTTTATGATTAATTCAGAGAGATCCGTTGAGAAAGAGACCAAC 1126  
 QY 2860 TATTAG 2865  
 DB 1127 TATTAG 1132  
 RESULT 12  
 AA160188  
 ID AA160188 standard; CDNA; 2209 BP.  
 AC AA160188;  
 XX 22-OCT-2001 (First entry)  
 DE Human polynucleotide SEQ ID NO 4177.  
 XX  
 KW Human; neotropic; immunosuppressant; cytostatic; gene therapy; cancer;  
 KW peripheral nervous system; neuropathy; central nervous system; CNS;  
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
 KW leukemia; ss.  
 XX  
 OS Homo sapiens.  
 PD W020015312-A1.  
 XX 26-JUL-2001.  
 PF 26-DEC-2000; 2000WO-US34263.  
 XX 21-JAN-2000; 2000US-0488725.  
 PR 25-APR-2000; 2000US-0552317.  
 PR 09-JUL-2000; 2000US-0598042.  
 PR 19-JUL-2000; 2000US-0620325.

```

PR 07-SEP-2000: 2000US-231167P.
XX
XX (INCY-) INCYTE GENOMICS INC.
XX
XX Jackson S., Lincoln SE, Altus CM, Dufour GE, Chalup MS, Hillman JL;
XX Jones AL, Yu JY, Wright RJ, Gietzen D, Liu TP, Yap PE, Dahl CR;
XX Momiyama MG, Bradley DL, Ronatski SP, Harris B, Roseberry AM;
XX Gerslun EH, Peralta CH, David MH, Panzer SR, Flores V, Delfo A;
XX Marwaha R, Chen AJ, Chang SC, Au AP, Imman RR;
XX
XX WPI: 2002-527544/56.
XX P-PSDB: ABP51423.
XX
XX Novel human disease detection and treatment polypeptide, useful in
XX diagnosis, prevention or treatment of cell proliferative disorders e.g.
XX arteriosclerosis, cirrhosis and an autoimmune/inflammatory disorder
XX e.g. AIDS
XX
XX Claim 1, Page 409-410; 61app; English.
XX
XX The invention relates to an isolated human disease detection and
XX treatment (MDOT) polypeptide (I) selected from a polypeptide having a
XX sequence selected from 254 sequences (ABP51231-ABP51484) given in the
XX specification, a naturally occurring polypeptide comprising a sequence
XX having at least 90% identity to (I) or a biologically active or
XX immunogenic fragment of (I). (I) is useful for screening a compound for
XX effectiveness as an agonist or antagonist, for screening a compound that
XX specifically binds (I) or modulates the activity of (I), and for
XX preparing a polyclonal or monoclonal antibody by hybridoma technology.
XX Nucleic acids (II) (ABQ12449-ABQ12700) encoding (I) are useful for
XX screening a compound for effectiveness in altering expression of a target
XX polynucleotide comprising. Oligonucleotides and antibodies are useful for
XX detecting MDOT in a sample or for assessing toxicity of a test compound,
XX in a diagnostic test for a condition or a disease associated with the
XX expression of MDOT in a biological sample, for detecting (I) in a sample,
XX and for purifying (I) from a sample. A composition comprising (I), an
XX agonist or antagonist is useful for treating a disease or condition.
XX (I) or (II) are useful for diagnosing, treating or preventing disorders
XX associated with decreased or increased expression of functional MDOT.
XX selected from a cell proliferative disorder such as arteriosclerosis,
XX cirrhosis, hepatitis, psoriasis, and cancer and an
XX autoimmune/inflammatory disorder such as AIDS, Addison's disease,
XX allergy, anaemia, asthma, atherosclerosis, gout, multiple sclerosis or
XX rheumatoid arthritis. (II) are useful for creating knockin humanised
XX animals or transgenic animals to model human diseases. In somatic or
XX germline gene therapy, to generate a transcript image of a tissue or cell
XX type, for detecting differences in the chromosomal location due to
XX translocation or inversion among normal, carrier or affected individuals
XX and as hybridisation probes for mapping naturally occurring genomic
XX sequences.
XX
XX Sequence 1184 BP; 368 A; 227 C; 272 G; 317 T; 0 other;
XX
XX Query Match 34.0%; Score 974; DB 24; Length 1184;
XX Best Local Similarity 100.0%; Pred. No. 0;
XX Matches 974; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 ATGGCTCACTATATACATTTCTCGCANGGTTTGGGCTGCTTCCAGATTCCTG 60
XX |||||
XX Db 211 ATGGCTCACTATATATACATTTCTCGCANGGTTTGGGCTGCTTCCAGATTCCTG 270
XX
XX 61 TTAGCTGAAGATGGGGAAGTAAAGATCAAGTGTCTACTGCTCCAGATTTAGTTTC 120
XX |||||
XX Db 271 TTAGCTGAAGATGGGGAAGTAAAGATCAAGTGTCTACTGCTCCAGATTTAGTTTC 330
XX
XX 121 ATCTTAGATGGCTCTTATAGTGTGGCCCAAGAACTTTGAAATAGTAAATAAGTGCCT 180
XX |||||
XX Db 331 ATCTTAGATGGCTCTTATAGTGTGGCCCAAGAACTTTGAAATAGTAAATAAGTGCCT 390
XX
XX 181 GTCAATATCCAAAAAATCTTGACATAGGGCCGAAAGTTATTTCAAGTTGAGTGTCAA 240
XX |||||
XX Db 391 GTCAATATCCAAAAAATCTTGACATAGGGCCGAAAGTTATTTCAAGTTGAGTGTCAA 450

```

```

Qy 241 TATAGTACTACCCCTGTGCTGAGATTCCTCTCGGAAGCTATGATTTGAGAGAACATTTG 300
|||
Db 451 TATAGTACTACCCCTGTGCTGAGATTCCTCTCGGAAGCTATGATTTGAGAGAACATTTG 510
Qy 301 ACCGAGAGAGTGGAAATCCATCTTACTTATAGAGAAACACAAAGACAGGAGGCCATC 360
|||
Db 511 ACCGAGAGAGTGGAAATCCATCTTACTTATAGAGAAACACAAAGACAGGAGGCCATC 570
Qy 361 CAGTTTGGCTGATTAACCTTTTGGCAAGTCCTCAGCATTTCTGACTTAAGATAGCATGTG 420
|||
Db 571 CAGTTTGGCTGATTAACCTTTTGGCAAGTCCTCAGCATTTCTGACTTAAGATAGCATGTG 630
Qy 421 GTACTTACGATGGCAATTCAGATGACGTCAAGATGACGTCAAGACAGACAT 480
|||
Db 631 GTACTTACGATGGCAATTCAGATGACGTCAAGATGACGTCAAGACAGACAT 690
Qy 481 AGTAGATTAACATTTATTTGCTATTTGTTGGTTCAGAAACAGAAAGATGCCAACTTGA 540
|||
Db 691 AGTAGATTAACATTTATTTGCTATTTGTTGGTTCAGAAACAGAAAGATGCCAACTTGA 750
Qy 541 GCTATTGGCAACAAAGCCTTCTGCTACTTATGTTTATGTGAAAGCTATTTGCAATA 600
|||
Db 751 GCTATTGGCAACAAAGCCTTCTGCTACTTATGTTTATGTGAAAGCTATTTGCAATA 810
Qy 601 TCCAAATTAAGGAAAGTATGAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 660
|||
Db 811 TCCAAATTAAGGAAAGTATGAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 870
Qy 661 ATTCAGATGGCAAGCTGTGATGAAGGAGATTTGATATCTTTTATGATTTAGATGTAAT 720
|||
Db 871 ATTCAGATGGCAAGCTGTGATGAAGGAGATTTGATATCTTTTATGATTTAGATGTAAT 930
Qy 721 AAAAAGTTAGAAAGAAATACAGCTTTACCAAAAAAAGATTAAGATTAAGATTAACA 780
|||
Db 931 AAAAAGTTAGAAAGAAATACAGCTTTACCAAAAAAAGATTAAGATTAAGATTAACA 990
Qy 781 TCAAAAGTTGATTTATCGAATCTCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 840
|||
Db 991 TCAAAAGTTGATTTATCGAATCTCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1050
Qy 841 TATGATTTGTTGCTTACTCAAGATTTAAAGTCAAGAAATTTGGATTTATGAGAAATA 900
|||
Db 1051 TATGATTTGTTGCTTACTCAAGATTTAAAGTCAAGAAATTTGGATTTATGAGAAATA 1110
Qy 901 TTAACATTTGATGGAAGGCCAAGATTAAGCAATTTAAATTTGGTGTGCAAAATCTTA 960
|||
Db 1111 TTAACATTTGATGGAAGGCCAAGATTAAGCAATTTAAATTTGGTGTGCAAAATCTTA 1170
Qy 961 TTATTTACACACAC 974
|||
Db 1171 TTATTTACACACAC 1184

```

```

RESULT 11
AAKS3029
ID AAKS3029 standard, cDNA; 2209 BP.
XX
XX AAKS3029;
XX
XX 06-NOV-2001 (first entry)
XX
XX Human polynucleotide SEQ ID NO 2558.
XX
XX Human cytokine; cell proliferation; cell differentiation; gene therapy;
XX vaccine; peptide therapy; stem cell growth factor; hematopoiesis;
XX tissue growth factor; immunomodulatory; cancer; leukemia;
XX nervous system disorder; arthritis; inflammation; ss.
XX
XX Homo sapiens.
XX
XX MO200157190-A2.
XX

```



PI Tang YF, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;  
PI Zhao Q, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
PI Xue AJ, Yang Y, Wehrman T, Goodrich R, Goodrich R;  
XX  
XX MPI: 2001-476283/51.  
DR P-PSDB: AAM78912.

PT Nucleic acids encoding polypeptides with cytokine-like activities,  
PT useful in diagnosis and gene therapy -

PS Claim 1: Page 2132-2134; 6221pp; English.

XX The invention relates to polynucleotides (AAK51456-AAK51435) and the  
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activity/inhibitin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666  
CC (AAM80020) are omitted as the relevant pages from the sequence listing  
CC were missing at the time of publication.

XX Sequence 2230 BP; 679 A; 409 C; 550 G; 592 T; 0 other:

Query Match 37.1%; Score 1062; DB 22; Length 2230;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1062; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1804 GGAATCCGAGATTTCTCGAACCAGGATTAATGGGCCCAAAAGGAGAAATTTGGGCTT 1863  
DB 91 GGAATCCGAGATTTCTCGAACCAGGATTAATGGGCCCAAAAGGAGAAATTTGGGCTT 150  
QY 1864 CCAGGAGACAAAGAAAAAGAGAGCCCGAGGATGCTGTTAATGGAAGCAATGSC 1523  
DB 151 CCAGGAGACAAAGAAAAAGAGAGCCCGAGGATGCTGTTAATGGAAGCAATGSC 210  
QY 1924 TCACGAGCCGACCTCGAACCAGGATCTTAAGGAGCAAAAGTGAACCTGGAATTCAA 1983  
DB 211 TCACGAGCCGACCTCGAACCAGGATCTTAAGGAGCAAAAGTGAACCTGGAATTCAA 270  
QY 1984 GGGATGCTCGGGGCTTCTGGGCTCAAGGAGAACACAGAGCAAGGGTCCCGAGAGAA 2043  
DB 271 GGGATGCTCGGGGCTTCTGGGCTCAAGGAGAACACAGAGCAAGGGTCCCGAGAGAA 330  
QY 2044 CCAGGATACATGGGTTTACCCGAGTTCAAGAAAAAGGGGACAAAGAAATCAAGGT 2103  
DB 331 CCAGGATACATGGGTTTACCCGAGTTCAAGAAAAAGGGGACAAAGAAATCAAGGT 390  
QY 2104 GAAAAAGGATTCAGGGTCGAAAAAGGAGAAATGAGACAGAGGAAATCCAGGAGCAAC 2163  
DB 391 GAAAAAGGATTCAGGGTCGAAAAAGGAGAAATGAGACAGAGGAAATCCAGGAGCAAC 450  
QY 2164 GGAATTCAGAGCCATCATGTGTGCAAAAGAGAGAGAGTGAAGAGGAGAACTGTGTCT 2223  
DB 451 GGAATTCAGAGCCATCATGTGTGCAAAAGAGAGAGTGAAGAGGAGAACTGTGTCT 510  
QY 2224 CGAGGTGCTATGATGTAAGAAAGAGAAATCTGGGTGATGCTGGGCGCCCGAGGT 2283  
DB 511 CGAGGTGCTATGATGTAAGAAAGAGAAATCTGGGTGATGCTGGGCGCCCGAGGT 570  
QY 2284 CCTAAGGGGCAACTGGGGATTCAGAGTCTCAAGGAGCCCGAGGTTTGAATGGAGCCG 2343  
DB 571 CCTAAGGGGCAACTGGGGATTCAGAGTCTCAAGGAGCCCGAGGTTTGAATGGAGCCG 630  
QY 2344 GGAAGAGATTTGAGACAAATTTATGACAAAGTTTGCACAGATGTAATAGAGCCAG 2403  
DB 631 GGAAGAGATTTGAGACAAATTTATGACAAAGTTTGCACAGATGTAATAGAGCCAG 690

QY 2404 CTACAGCTTACTTCAGAGTGGAGAGATTGAAATTTGATCATTCGCTGCCAAT 2463  
DB 631 CTACAGCTTACTTCAGAGTGGAGAGATTGAAATTTGATCATTCGCTGCCAAT 750  
QY 2464 GGGTCCCCGGGATTTCTGGGCACTGGTCCGATAGGCCCGAGAGAGGTCCAGAGAGATTA 2523  
DB 751 GGGTCCCCGGGATTTCTGGGCACTGGTCCGATAGGCCCGAGAGAGGTCCAGAGAGATTA 810  
QY 2524 CCTGGTTGGCAGAGAGATGATGTTCTGAGATAGTGGTGTCCCTGGAGGTCCAGGT 2583  
DB 811 CCTGGTTGGCAGAGAGATGATGTTCTGAGATAGTGGTGTCCCTGGAGGTCCAGGT 870  
QY 2584 GTCAAGAGATTTAAAGGCTTACAGAGAAATGGGAGAAAGAGAGCCAAAGGTTTGGG 2643  
DB 871 GTCAAGAGATTTAAAGGCTTACAGAGAAATGGGAGAAAGAGAGCCAAAGGTTTGGG 930  
QY 2644 TATCTGGAGAGAGAGTCTGCTGGGCCCCCGAGGTCGAGAGGCCCCCTCTGGAATTAAG 2703  
DB 931 TATCTGGAGAGAGAGTCTGCTGGGCCCCCGAGGTCGAGAGGCCCCCTCTGGAATTAAG 990  
QY 2704 AAAGAGGTCCTCCAGAGAGACCAAGTCTCTGGCCAAAGATGGAACATGAGAAACCT 2763  
DB 991 AAAGAGGTCCTCCAGAGAGACCAAGTCTCTGGCCAAAGATGGAACATGAGAAACCT 1050  
QY 2764 GGAATCCAGAGGCAAGGAGCCCGCCAGGATCTGAGACCCATCACTATTTAGTGA 2823  
DB 1051 GGAATCCAGAGGCAAGGAGCCCGCCAGGATCTGAGACCCATCACTATTTAGTGA 1110  
QY 2824 ATTGCCAGAGAGATTCGTTTCAAGAAAAAGAGCAAACTATTAG 2865  
DB 1111 ATTGCCAGAGAGATTCGTTTCAAGAAAAAGAGCAAACTATTAG 1152

RESULT 9  
AA158402  
ID AA158402 standard; cDNA; 2230 BP.  
XX  
XX AA158402;  
XX  
CT 22-OCT-2001 (first entry)  
XX  
XX Human polynucleotide SEQ ID NO 605.  
DS  
XX Human; noctropic; immunosuppressant; cytostatic; gene therapy; cancer;  
XX peripheral nervous system; neuropathy; central nervous system; CNS;  
XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemostatic;  
XX chemokinetic; tymolytic; drug screening; arthritis; inflammation;  
XX leukaemia; ss.  
XX  
OS Homo sapiens.  
XX  
PN W0200153312-A1.  
XX  
XX 26-JUL-2001.  
XX  
XX 26-DEC-2003; 2000US-0534263.  
XX  
XX 21-JAN-2003; 2000US-0488725.  
XX  
XX 25-APR-2003; 2000US-0552317.  
XX  
XX 09-JUL-2000; 2000US-0596042.  
XX  
XX 19-JUL-2000; 2000US-0620312.  
XX  
XX 03-AUG-2000; 2000US-0653450.  
XX  
XX 14-SEP-2000; 2000US-0662191.  
XX  
XX 19-OCT-2000; 2000US-0693036.  
XX  
XX 29-NOV-2000; 2000US-0727344.  
XX  
XX  
XX (HENSE) HENSE INC.  
XX  
XX Tang YF, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
XX Zhao Q, Zhou P, Goodrich R, Drmanac RT;

CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activin/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukemia, nervous system disorders, arthritis and  
 CC inflammation.  
 CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666  
 CC (AAK80020) were omitted as the relevant pages from the sequence listing  
 CC were missing at the time of publication.

XX Sequence 1691 BP, 532 A; 326 C; 386 G; 447 T; 0 other;

Query Match 42.0%; Score 1203; DB 22; Length 1691;  
 Best Local Similarity 99.8%; Pred. No. 0;

Matches 1423; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 ATGGCTACATATATTAATTTCTGTGATGGTTTGGTGTGCTTCTTCAGAAATCTGTG 60  
 Db 115 ATGGCTACATATTAATTTCTGTGATGGTTTGGTGTGCTTCTTCAGAAATCTGTG 174  
 QY 61 TTACCTGAAGATGGGGAAGTAAATCAAGTTTGTCTGCTCCGACAGATTAGTTTC 120  
 Db 175 TTACCTGAAGATGGGGAAGTAAATCAAGTTTGTCTGCTCCGACAGATTAGTTTC 234  
 QY 121 ATCTTAGATGGCTCTTATAGTGTGGCCAGAAAACCTTGAAAATAGTGAAGAGTGGCTT 180  
 Db 235 ATCTTAGATGGCTCTTATAGTGTGGCCAGAAAACCTTGAAAATAGTGAAGAGTGGCTT 234  
 QY 181 GTCAATATCAAAAAAATTGACATAGGCGCCGAAGTTTATCAAGTTGAGTGTCAA 240  
 Db 295 GTCAATATCAAAAAAATTGACATAGGCGCCGAAGTTTATCAAGTTGAGTGTCAA 354  
 QY 241 TATAGTGAATACCTGTGTGGAGATTCCTCTGCGAAGCTATGATTCAGAGAAACATTG 300  
 Db 355 TATAGTGAATACCTGTGTGGAGATTCCTCTGCGAAGCTATGATTCAGAGAAACATTG 414  
 QY 301 ACGGACAGATGGAATCCATATCTTACTTAGAGAGAAACA CAAGA CAGGAAAGCCATC 360  
 Db 415 ACGGACAGATGGAATCCATATCTTACTTAGAGAGAAACA CAAGA CAGGAAAGCCATC 474  
 QY 361 CAGTTGGCTCGGATTCCTTTTGGCAAGTCCGACAGATTTGTGATAGATAGCAGT 420  
 Db 475 CAGTTGGCTCGGATTCCTTTTGGCAAGTCCGACAGATTTGTGATAGATAGCAGT 534  
 QY 421 GTACTTAGATGGAATCCCAAGATGACGTCAAGATGACGTCAAGCAGCAGAGAT 480  
 Db 535 GTACTTAGATGGAATCCCAAGATGACGTCAAGATGACGTCAAGCAGCAGAGAT 594  
 QY 481 AGTAAATTAACATTTTGTCTATTGTGTGTTGAGAAACAGAAATGCGCAACTTGA 540  
 Db 595 AGTAAATTAACATTTTGTCTATTGTGTGTTGAGAAACAGAAATGCGCAACTTGA 654  
 QY 541 GCTATGCGCAACAGCCCTGCTGCTACTTATGTTTATGAGAAAGCTATATTTGAA 600  
 Db 655 GCTATGCGCAACAGCCCTGCTGCTACTTATGTTTATGAGAAAGCTATATTTGAA 714  
 QY 601 TCCAAAATTAAGGAAGTATGAAAGCAGAACTTGTGAGAAATCTGTCTCCACAGCA 660  
 Db 715 TCCAAAATTAAGGAAGTATGAAAGCAGAACTTGTGAGAAATCTGTCTCCACAGCA 774  
 QY 661 ATTCCAGTGGAGCTGTGTGATGAAAGGGGATTTGATTTCTTTAGGTTAGTGTAAAT 720  
 Db 775 ATTCCAGTGGAGCTGTGTGATGAAAGGGGATTTGATTTCTTTAGGTTAGTGTAAAT 834  
 QY 721 AAAAAGTTAAGAAAGATACAGCTTTCACAAAAGATTAAGAGATATGAAATGACA 780  
 Db 835 AAAAAGTTAAGAAAGATACAGCTTTCACAAAAGATTAAGAGATATGAAATGACA 854  
 QY 781 TCAAAAGTTGATTTATAGAACTCAACAAGATGTTTTCCCAAGAGTCTTCTCCATCA 840  
 Db 895 TCAAAAGTTGATTTATAGAACTCAACAAGATGTTTTCCCAAGAGTCTTCTCCATCA 954  
 QY 841 TATGATTTGTGTACTCAAAAGTTTAAAGTCAAGAAATTTGGAGTTATAGAAATA 900

Db 955 TATGATTTGTGTCTACTCAAAAGTTTAAAGTCAAGAAATTTGGGATTTTAGAGATA 1014  
 QY 901 TTAACATATTATGGAAGGCCCAATATGACAGTTACCTTAAATGGTGTGACAAATCTTA 960  
 Db 1015 TTAACATATTATGAG-ATGCCCAAAATGACAGTTACCTTAAATGGTGTGACAAATCTTA 1073  
 QY 961 TTATTTCAAAACACAGCGTATTTAATGAGTGTGACAGATGAGTGTACCTTGTACCCGCA 1020  
 Db 1074 TTATTTCAAAACACAGCGTATTTAATGAGTGTGACAGATGAGTGTACCTTGTACCCGCA 1133  
 QY 1021 GTTAAAGCTTGTGTATGATGACCAACCAATTTGAAAGAGCCCTTACATCCAGTTTAAAG 1080  
 Db 1134 GTTAAAGCTTGTGTATGATGACCAACCAATTTGAAAGAGCCCTTACATCCAGTTTAAAG 1193  
 QY 1141 ATCTTAGATGATGGGCAAAACCCCAATTTGAAAGATATCTGTGAAAGAAAGAACTGTGAG 1200  
 Db 1254 ATCTTAGATGATGGGCAAAACCCCAATTTGAAAGATATCTGTGAAAGAAAGAACTGTGAG 1313  
 QY 1201 TTGATGTCCAAAAGTTGCGAATCTACTGTGACCCAGAAAGCAACAACCGGAGACAGCA 1260  
 Db 1314 TTGATGTCCAAAAGTTGCGAATCTACTGTGACCCAGAAAGCAACAACCGGAGACAGCA 1373  
 QY 1261 TGTGAGATTTCTGATTTTGGCTTAAATGATCCAGTATGTAGTTCACTCCAGCTTCC 1320  
 Db 1374 TGTGAGATTTCTGATTTTGGCTTAAATGATCCAGTATGTAGTTCACTCCAGCTTCC 1433  
 QY 1321 TGTATTTGTCTCCGGGAAACCAAGAGCTTCAAGGCCCAAGAGTGAAGCTGAGCTGCT 1380  
 Db 1434 TGTATTTGTCTCCGGGAAACCAAGAGCTTCAAGGCCCAAGAGTGAAGCTGAGCTGCT 1493  
 QY 1381 GGGAAACCTGTGCTACCTGTGACAACTGTGTGAAGATGTAAAGCTG 1426  
 Db 1494 GGGAAACCTGTGCTACCTGTGACAACTGTGTGAAGATGTAAAGCTG 1539

RESULT 8  
 ID AAK52045  
 AAK52045 standard; cDNA; 2230 BP.  
 XX  
 AC AAK52045;  
 XX  
 DT 06-NOV-2001 (first entry)  
 XX  
 DE Human polynucleotide SEQ ID NO 590.  
 XX  
 KW Human; dykine; cell proliferation; cell differentiation; gene therapy;  
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 KW tissue growth factor; immunomodulatory; cancer; leukemia;  
 KW nervous system disorder; arthritis; inflammation; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN MO20157:90-42.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 05-FEB-2001; 2001MC-0504098.  
 XX  
 PR 03-FEB-2003; 2000US-0496914.  
 PR 27-APR-2003; 2000US-0560875.  
 PR 20-JUN-2000; 2000US-0598375.  
 PR 19-JUL-2000; 2000US-0620325.  
 PR 31-SEP-2000; 2000US-0634936.  
 PR 15-SEP-2000; 2000US-0635361.  
 PR 20-OCT-2000; 2000US-0633125.  
 PR 30-NOV-2000; 2000US-0728422.  
 XX  
 PA HYSEQ INC.  
 XX

QY	6	TTAGCTGAAGATGGGGAAGTAGAATCAAGTTGTGTACTGCTCCGACACATTTACTTTTC	120
Db	62	TTAGCTGAAGATGGGGAAGTAGAATCAAGTTGTGTACTGCTCCGACACATTTACTTTTC	121
QY	121	ATCTTAGATGGCTTTATATGTGTGGCCACAGAAAATTGAAATAGTGA AAAATGGCGTT	180
Db	122	ATCTTAGATGGCTTTATAGTGTGGCCACAGAAAATTGAAATAGTGA AAAATGGCGTT	181
QY	181	GTCAAATATCAAAAAAATCTTGACATGAGGCGAAAGTTATTAAGCTTGAGTGGTCAA	240
Db	182	GTCAAATATCAAAAAAATCTTGACATGAGGCGAAAGTTATTAAGCTTGAGTGGTCAA	241
QY	241	TATAGTGAATACCTGTGTGTGAGATTCCTTCGGAAGCTATGATATTCAGAGACATTTG	300
Db	242	TATAGTGAATACCTGTGTGTGAGATTCCTTCGGAAGCTATGATATTCAGAGACATTTG	301
QY	301	ACGGACAGAGTGGAAATCCATCTACTTATGAGAAACCAAGACAGAGAAAGCCATTC	360
Db	302	ACGGACAGAGTGGAAATCCATCTACTTATGAGAAACCAAGACAGAGAAAGCCATTC	361
QY	361	CAGTTTGGCGTGCATTAACCTTTTGGCAAGTCTCAGATTCTGACTAAGATAGCAGTG	420
Db	362	CAGTTTGGCGTGCATTAACCTTTTGGCAAGTCTCAGATTCTGACTAAGATAGCAGTG	421
QY	421	GTAAGTACGGATGGCAAAATCCCAAGATACGTCAGAGATCAGCTCAACAGACAGAGAT	480
Db	422	GTAAGTACGGATGGCAAAATCCCAAGATACGTCAGAGATCAGCTCAACAGACAGAGAT	481
QY	481	AGTAAGATTAACATTAATTGTCTATTTGGTGTGTGTTCAGAAAACAGAAATGCGCACTTGA	540
Db	482	AGTAAGATTAACATTAATTGTCTATTTGGTGTGTGTTCAGAAAACAGAAATGCGCACTTGA	541
QY	541	GCTATTGCGCAACAGGCTTCGTCTACTATGTGTTTATGTGGAAAGATATATTCGATA	600
Db	542	GCTATTGCGCAACAGGCTTCGTCTACTATGTGTTTATGTGGAAAGATATATTCGATA	601
QY	601	TCCAAATATANGGAAGTGTATGAAGCAAACTTTGTGAAGATGTGTGTGTCCAAACAGA	660
Db	602	TCCAAATATANGGAAGTGTATGAAGCAAACTTTGTGAAGATGTGTGTGTCCAAACAGA	661
QY	661	ATTCCAGTGGACAGCTGCTATGATAAAGGGAATTTGATATCTTTTAGGTTAGATGAAT	720
Db	662	ATTCCAGTGGACAGCTGCTATGATAAAGGGAATTTGATATCTTTTAGGTTAGATGAAT	721
QY	721	AAAAAGTTAAGAAAAGATACAGCTTTCACCAAAAAGATAAAGGATATGAGATACAA	780
Db	722	AAAAAGTTAAGAAAAGATACAGCTTTCACCAAAAAGATAAAGGATATGAGATACAA	781
QY	781	TCAAAAGTTAATATCAGAACTCACAAGCAATGTTTCCAGAAAGTCTTCCTCATCA	840
Db	782	TCAAAAGTTAATATCAGAACTCACAAGCAATGTTTCCAGAAAGTCTTCCTCATCA	841
QY	841	TATGTATTGTGTCTACTCAAGATTTTAAAGTCAAGAAAATTGGGATTTATGAGAGATA	900
Db	842	TATGTATTGTGTCTACTCAAGATTTTAAAGTCAAGAAAATTGGGATTTATGAGAGATA	901
QY	901	TTAAGTATTGATGGAAGGCGACAGAAATAGCAGTTACCTTAATATGGTGTGCAAAATCTTA	960
Db	902	TTAAGTATTGATGGAAGGCGACAGAAATAGCAGTTACCTTAATATGGTGTGCAAAATCTTA	961
QY	961	TTATTATCAACAACGAGCTAATTAATGCTTCACAAAGTGTACCTTTGGTAAACCTGAA	1020
Db	961	TTATTATCAACAACGAGCTAATTAATGCTTCACAAAGTGTACCTTTGGTAAACCTGAA	1021
QY	1021	GTTAAGACGTTGTTTATGATGAAGGCTGCGACCAAAATCGTCTCTTAGTAAACAGAACAGAT	1080
Db	1021	GTTAAGACGTTGTTTATGATGAAGGCTGCGACCAAAATCGTCTCTTAGTAAACAGAACAGAT	1081
QY	1081	GTCAGCTTGATATTTGATGACCAACAAATGTA AAAACAGCCCTTACATCAGTTTAGGG	1140
Db	1081	GTCAGCTTGATATTTGATGACCAACAAATGTA AAAACAGCCCTTACATCAGTTTAGGG	1141

Cy	1144	ATCTGATCAAGGGGCAAAACCCAAATGTGAAAAATATTCTGGAAAAAGAAACCTTCAG	1200
Ds	1141	ATCTTGATCAAGGGGCAAAACCCAAATGTGAAAAATATTCTGGAAAAAGAAACCTTCAG	1200
Cy	1201	TTTGATGTCGAAAAGTTGGCGAATCTACTGTGACCCGAAACAGAAACAAACCCGAGACAGCA	1260
Ds	1201	TTTGATGTCGAAAAGTTGGCGAATCTACTGTGACCCGAAACAGAAACAAACCCGAGACAGCA	1260
Cy	1261	TGTGGAGATCCCTCGAATTGGCTTAATGTGCCAGATGATGAGTCAACCTCAAGCTCC	1320
Ds	1261	TGTGGAGATCCCTCGAATTGGCTTAATGTGCCAGATGATGAGTCAACCTCAAGCTCC	1320
Cy	1321	TGTATTGTTCCTCCGGGAAAAACAGAGCTTCAAGGCCCCAAAGGTGACCTCTGACCTGCT	1380
Ds	1321	TGTATTGTTCCTCCGGGAAAAACAGAGCTTCAAGGCCCCAAAGGTGACCTCTGACCTGCT	1380
Cy	1381	GGGAAACCTGTGCTACCTCGGCACACACTGGTCAAGATGGATTAAGCTG	1426
Ds	1381	GGGAAACCTGTGCTACCTCGGCACACACTGGTCAAGATGGATTAAGCTG	1426

## RESULT

ID AAK52265 standard; cDNA; 1691 bp

AC AAK52265;

DT 06-NOV-2001 (first entry)

DE Human polynucleotide SEQ ID NO 810.

KW :human; cytokine; cell proliferation; cell differentiation; gene therapy;

KW tissue growth factor; immunomodulatory; cancer; leukemia;

XX

XXI

XX	3
XX	4
XX	5
XX	6
XX	7
XX	8
XX	9
XX	10
XX	11
XX	12
XX	13
XX	14
XX	15
XX	16
XX	17
XX	18
XX	19
XX	20
XX	21
XX	22
XX	23
XX	24
XX	25
XX	26
XX	27
XX	28
XX	29
XX	30
XX	31
XX	32
XX	33
XX	34
XX	35
XX	36
XX	37
XX	38
XX	39
XX	40
XX	41
XX	42
XX	43
XX	44
XX	45
XX	46
XX	47
XX	48
XX	49
XX	50
XX	51
XX	52
XX	53
XX	54
XX	55
XX	56
XX	57
XX	58
XX	59
XX	60
XX	61
XX	62
XX	63
XX	64
XX	65
XX	66
XX	67
XX	68
XX	69
XX	70
XX	71
XX	72
XX	73
XX	74
XX	75
XX	76
XX	77
XX	78
XX	79
XX	80
XX	81
XX	82
XX	83
XX	84
XX	85
XX	86
XX	87
XX	88
XX	89
XX	90
XX	91
XX	92
XX	93
XX	94
XX	95
XX	96
XX	97
XX	98
XX	99
XX	100

	6	7	8	9	0
XX					

[illegible]

27-APR-2000; 2000US-0560875

PR 20-JUL-2000; 2000US-D-35807, 2  
19-JUL-2000; 2000US-3620325.

01-SEP-2000: 2000US-0663561

PR 20-CC1-2000; 2000US-0633323  
30-NOV-2000; 2000US-0728422

xx  
DA (HYSE-) HYSEO INC.

xx  
PI  
Tand VT  
Till C  
Dymanac RT

PI Zhaod Qia, wang D, wang J,  
Xue Ai, yang Y, Wei-hyman

XX  
DB  
WPT: 2001-476283/51

DR P-PSJB; AAM/9132.  
XX

**Nucleic acids encoding polymers useful in diagnosis and gene**

XX  
PS  
CJajm 1: Page 2710-2711: 622

xx The invention relates to pol  
xx

CC encoded polypeptides (AA1/8.1  
CC cytokine cel) proinflammatory

production of other cytokine

CC peptide therapy. The polypeptide



```

Db      1381 GGGAACTCTGCTACCTGGACACCTGCTCAAGATGTAAGCTTG 1426
302 ACCGACAGATGGATCCATCTCTACTTAGAGAGAAACAAAGCAGGAGAGCCCTC 361
361 CAGTTTGGCTGCGATTACCTTTTGGCCAGTCTCTCGATTTCTGACTAAGATAGCAGT 420
362 CAGTTTGGCTGCGATTACCTTTTGGCCAGTCTCTCGATTTCTGACTAAGATAGCAGT 421
421 GTACTTAAGATGGCAATCCCAAGATGACGTGACAGCTCAGCAGCAGAGAGAT 480
422 GTACTTAAGATGGCAATCCCAAGATGACGTGACAGCTCAGCAGCAGAGAT 481
481 AGTAAATACATATTCTGATTTGGTGTGTCGAAACGAAATGCGAATCTTGA 540
482 AGTAAATACATATTCTGATTTGGTGTGTCGAAACGAAATGCGAATCTTGA 541
541 GCTATTGCGCAACAGCTCTGCTACTTATGTTTATGTGGAGACTATTTGCAATA 600
542 GCTATTGCGCAACAGCTCTGCTACTTATGTTTATGTGGAGACTATTTGCAATA 601
601 TCCAAATTAAGGAGATGATGAGAGCAAACTTTGAGAAATCTGTCTGCAACAGCA 660
602 TCCAAATTAAGGAGATGATGAGAGCAAACTTTGAGAAATCTGTCTGCAACAGCA 661
661 ATTCAGTGGCAGCTCTGATGAGAGGGGATTTGATATTCTTTAGGTTAGATGTAAT 720
662 ATTCAGTGGCAGCTCTGATGAGAGGGGATTTGATATTCTTTAGGTTAGATGTAAT 721
721 AAAAAGGTTAAGAAAGATACAGCTTCCACAAAAGATTAAGAGATATGAAATACA 780
722 AAAAAGGTTAAGAAAGATACAGCTTCCACAAAAGATTAAGAGATATGAAATACA 781
781 TCAAAAGTATTTATGAGAACTCACAGCAATGTTTCCAGAGAGTCTTCTCATCA 840
782 TCAAAAGTATTTATGAGAACTCACAGCAATGTTTCCAGAGAGTCTTCTCATCA 841
841 TATGATTTGTGTCTACTCAAAAGATTTAAAGTCAAGAAATTTGCGATTTATGAGATA 900
842 TATGATTTGTGTCTACTCAAAAGATTTAAAGTCAAGAAATTTGCGATTTATGAGATA 901
901 TTAAGTATTTGATGAGAGGCAAAATAGCAGTTACCTTAATGTTGAGCAAAATCTTA 960
902 TTAAGTATTTGATGAGAGGCAAAATAGCAGTTACCTTAATGTTGAGCAAAATCTTA 961
961 TTAATTAACAAACACAGCTATTAATGAGTCAACAGTGTACTCCTTAACAGAAACAGAT 1020
962 TTAATTAACAAACACAGCTATTAATGAGTCAACAGTGTACTCCTTAACAGAAACAGAT 1021
1021 GTTAAGACCTTTGATGAGAGGCTGACCAAAATCTCTTACTTAACAGAAACAGAT 1080
1022 GTTAAGACCTTTGATGAGAGGCTGACCAAAATCTCTTACTTAACAGAAACAGAT 1081
1081 GTGACTTTGATGATGAGACCAAAATTTGAAACAGAGCCCTTAATCAGTTTAAAGG 1140
1082 GTGACTTTGATGATGAGACCAAAATTTGAAACAGAGCCCTTAATCAGTTTAAAGG 1141
1141 ATCTTGATCAATGAGGCAAAACCAAAATTTGAAACAGAGCCCTTAATCAGTTTAAAGG 1200
1142 ATCTTGATCAATGAGGCAAAACCAAAATTTGAAACAGAGCCCTTAATCAGTTTAAAGG 1201
1201 TTTGATGTCGCAAAAGTTGCGAAATCTACTGAGACCCAGAAACAGAAACCGGAGACAGCA 1260
1202 TTTGATGTCGCAAAAGTTGCGAAATCTACTGAGACCCAGAAACAGAAACCGGAGACAGCA 1261
1261 TGTGAGATCTCTGATTTTGTCTTAATGTCCTCAAGTATGATGATTAATCAGTCTCC 1320
1262 TGTGAGATCTCTGATTTTGTCTTAATGTCCTCAAGTATGATGATTAATCAGTCTCC 1321
1321 TGTGATTTGCTCTCGGAGAAACAGGAGCTTCAAGAGCCCAAGGTTGAGCTGAGCTGCT 1380
1322 TGTGATTTGCTCTCGGAGAAACAGGAGCTTCAAGAGCCCAAGGTTGAGCTGAGCTGCT 1381
1381 GGGAACTCTGCTACCTGGACACCTGCTCAAGATGTAAGCTTG 1426

```

```

Db      1381 GGGAACTCTGCTACCTGGACACCTGCTCAAGATGTAAGCTTG 1426
RESULT 6
AAK53249
ID      AAK53249 standard; cDNA; 1578 bp.
XX
XX      AAK53249;
XX
XX      06-NOV-2001 (first entry)
XX
DE      Human polynucleotide SEQ ID NO 2778.
XX
XX      Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX      vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX      tissue growth factor; immunomodulatory; cancer; leukaemia;
XX      nervous system disorder; arthritis; inflammation; ss.
OS      Homo sapiens.
XX
XX      W0200157190-A2.
XX
XX      09-AUG-2001.
XX
XX      05-FEB-2001; 2001WO-US04098.
XX
XX      03-FEB-2000; 2000US-0496914.
XX      27-APR-2000; 2000US-0560875.
XX      19-JUL-2000; 2000US-0598075.
XX      15-SEP-2000; 2000US-0620325.
XX      15-SEP-2000; 2000US-0654936.
XX      20-OCT-2000; 2000US-0693325.
XX      30-NOV-2000; 2000US-0728422.
XX
XX      (HYSE-) HYSEQ INC.
XX
XX      Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
XX      Zhao GA, Wang Y, Zhang J, Ren F, Chen R, Wang ZW;
XX      Xue Ai, Yang Y, Wejhrman T, Goodrich R;
XX      WP1: 2001-476283/51.
XX      P-PSDS; AAM80116.
XX
XX      Nucleic acids encoding polypeptides with cytokine-like activities,
XX      useful in diagnosis and gene therapy.
XX
XX      Claim 1; Page 4980-4981; 6221pp; English.
XX
XX      The invention relates to polynucleotides (AAK51456-AAK53435) and the
XX      encoded polypeptides (AAM78323-AAM80302) that exhibit activity relating to
XX      cytokine, cell proliferation or cell differentiation or which may induce
XX      production of other cytokines in other cell populations. The
XX      polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX      peptide therapy. The polypeptides have various cytokine-like activities,
XX      e.g. stem cell growth factor activity, haematopoiesis regulating
XX      activity, tissue growth factor activity, immunomodulatory activity and
XX      activity/inhibin activity and may be useful in the diagnosis and/or
XX      treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX      inflammation.
XX      Note: Records for SEQ ID NO 2:10 (AAK52582), 211; (AAK52582) and 3666
XX      (AAM80020) are omitted as the relevant pages from the sequence listing
XX      were missing at the time of publication.
XX
XX      Sequence 1578 BP; 496 A; 294 C; 359 G; 429 T; 0 other;
XX
XX      Query Match      43.8%; Score 1254; DB 22; Length 1578;
XX      Best Local Similarity 59.9%; Pred. No. 0;
XX      Matches 1424; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

```



```

QY 2064 CGGGAATTCAGGAAAAAGGAGGAGCAAAAGAAATCAAGTGAAGAAAAAGTATTCAGGATCA 2:23
DB 489 CGGGAATTCAGGAAAAAGGAGGAGCAAAAGAAATCAAGTGAAGAAAAAGTATTCAGGATCA 5:48
QY 2124 AAAAGGAGAAAAATGAAAGCAGGAAATTCAGAGGCAACAGAGAAATTCAGAGCCATCATGG 2:183
DB 549 AAAAGGAGAAAAATGAAAGCAGGAAATTCAGAGGCAACAGAGAAATTCAGAGCCATCATGG 6:08
QY 2184 TCGAAAGAGAGAGAGGATGAAAAAGGAGAACTGGTCTCCGAGGTGCATTTGGATCAAA 22:43
DB 609 TCGAAAGAGAGAGAGGATGAAAAAGGAGAACTGGTCTCCGAGGTGCATTTGGATCAAA 66:8
QY 2244 AGGAGAAATCTGGAGTGGATGGCTTGAATGGAGCCCGCAGAGTCTTAAGGGGCAACCTGGGGA 23:03
DB 669 AGGAGAAATCTGGAGTGGATGGCTTGAATGGAGCCCGCAGAGTCTTAAGGGGCAACCTGGGGA 72:8
QY 2304 TCCAGGTCTCAGGAGACCCCAAGGTTTGGATGGGAAGCCGGAGAGAGATTTTCAGAGCA 23:63
DB 729 TCCAGGTCTCAGGAGACCCCAAGGTTTGGATGGGAAGCCGGAGAGAGATTTTCAGAGCA 78:8
QY 2364 ATTATTCAGCAAGTTTGCAGAGATGATTAAGAGCCAGCTTACAGTCTTACTTCAGAG 24:23
DB 789 ATTATTCAGCAAGTTTGCAGAGATGATTAAGAGCCAGCTTACAGTCTTACTTCAGAG 84:8
QY 2424 TGGAAAGAAATGAAATTTGTGATTCGCTCTCCACATGGCTCCCGGGTATTCCTGG 24:83
DB 849 TGGAAAGAAATGAAATTTGTGATTCGCTCTCCACATGGCTCCCGGGTATTCCTGG 9:8
QY 2484 GCCACCTGGTCCGATAGGCCCAAGAGGTCCTCCAGAGATTAAGTCTTGGCAGAGAGAG 25:43
DB 909 GCCACCTGGTCCGATAGGCCCAAGAGGTCCTCCAGAGATTAAGTCTTGGCAGAGAGAG 96:8
QY 2544 TGGTGTCTCTGATTAAGTGGGTGTCTCTCGACGTCAGAGTGTCAAGAGATTTAAAGGCTT 26:03
DB 969 TGGTGTCTCTGATTAAGTGGGTGTCTCTCGACGTCAGAGTGTCAAGAGATTTAAAGGCTT 102:8
QY 2604 ACCAGGAAAGAAATGGGAGAAAAAGGAGCAAGGCTTGGGTATCTGGAAGAACAGGCTC 26:53
DB 1029 ACCAGGAAAGAAATGGGAGAAAAAGGAGCAAGGCTTGGGTATCTGGAAGAACAGGCTC 108:8
QY 2664 TCGTGGTCTCCAGGTCAGAGGCTCTCTCGATGAATGAAGAAAGGTCCTCCAGAGAG 27:23
DB 1089 TCGTGGTCTCCAGGTCAGAGGCTCTCTCGATGAATGAAGAAAGGTCCTCCAGAGAG 114:8
QY 2724 CCGAGGTCTCTCGCAAGATGAGACATCGAAACCTGGATCCAGAGGCAACGAGG 27:83
DB 1149 CCGAGGTCTCTCGCAAGATGAGACATCGAAACCTGGATCCAGAGGCAACGAGG 120:8
QY 2784 CCCCCAGGATCTCGAGCCCATCTACTATGTTTAAATTCAGAGAGAGATTCCTG 28:43
DB 1209 CCCCCAGGATCTCGAGCCCATCTACTATGTTTAAATTCAGAGAGAGATTCCTG 126:8
QY 2844 CAGAAAGAGACCAAACTATTAG 28:65
DB 1269 CAGAAAGAGACCAAACTATTAG 129:0

```

## RESULT 5

ID AAK53227 standard; cDNA; 1578 BP.

AAK53227;

06-NOV-2001 (first entry)

Human polynucleotide SEQ ID NO 2756.

Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 tissue growth factor; immunomodulatory; cancer; leukemia;  
 nervous system disorder; arthritis; inflammation; ss.

```

OS Homo sapiens.
XX
XX WO200157190-A2.
XX
XX 09-AUG-2001.
XX
XX 05-FEB-2001; 2001WO-US04098.
XX
XX 03-FEB-2000; 2000US-0496914.
XX 27-APR-2000; 2000US-0560875.
XX 20-JUN-2000; 2000US-0598075.
XX 19-JUL-2000; 2000US-0620325.
XX 01-SEP-2000; 2000US-0654936.
XX 15-SEP-2000; 2000US-0663561.
XX 20-OCT-2000; 2000US-0693325.
XX 30-NOV-2000; 2000US-0728422.
XX
XX (HSE-) HYSEQ INC.
XX
XX Tang YF, Liu C, Dimauc RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
XX Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen S, Wang ZW;
XX Xue AJ, Yang Y, Wejman T, Goodrich R;
XX
XX WPI: 2006-476283/51.
XX P-PSDB: AAK80994.
XX
XX Nucleic acids encoding polypeptides with cytokine-like activities,
XX useful in diagnosis and gene therapy -
XX
XX Claim 1: Page 4966; 6221pp; English.
XX
XX The invention relates to polynucleotides (AAK51456-AAK51435) and the
XX encoded polypeptides (AAK78323-AAK80302) that exhibit activity relating to
XX cytokine, cell proliferation or cell differentiation or which may induce
XX production of other cytokines in other cell populations. The
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX peptide therapy. The polypeptides have various cytokine-like activities,
XX e.g. stem cell growth factor activity, haematopoiesis regulating
XX activity, tissue growth factor activity, immunomodulatory activity and
XX activity/inhibin activity and may be useful in the diagnosis and/or
XX treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX inflammation.
XX Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
XX (AAK80300) are omitted as the relevant pages from the sequence listing
XX were missing at the time of publication.
XX
XX Sequence 1578 BP; 496 A; 294 C; 359 G; 429 T; 0 other;
XX
XX Query Match 43.8%; Score 1254; DB 22; Length 1578;
XX Best Local Similarity 99.9%; Pred. No. 0;
XX Matches 1424; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
XX
QY 1 ATGGCTCACTAATTAATTAATTTCTCGATGGTGGTGGTCTTCAGAAATTCGTG 60
DB 2 ATGGCTCACTAATTAATTAATTTCTCGATGGTGGTGGTCTTCAGAAATTCGTG 61
QY 61 TTAGCTGAAGATGGGAGATGAATCAAGTGTCTGACTCTCCGACAGATTTAGTTTC 120
DB 62 TTAGCTGAAGATGGGAGATGAATCAAGTGTCTGACTCTCCGACAGATTTAGTTTC 121
QY 121 ATCTTAGATGGCTCTTAATAGTGTGGCCCAAAAACCTTGAATAGGAAAAAGTGGCT 180
DB 122 ATCTTAGATGGCTCTTAATAGTGTGGCCCAAAAACCTTGAATAGGAAAAAGTGGCT 181
QY 181 GTCAATATCACAAAAAAATTTGACATAGGGCCGAGTTTATCAAGTTGAGAGTTCAC 240
DB 182 GTCAATATCACAAAAAAATTTGACATAGGGCCGAGTTTATCAAGTTGAGAGTTCAC 241
QY 241 TATAGTACTACCTGTGCTGAGATTCCTCTCGAAAGCTATGATTCAGAGAAACATTTG 300
DB 242 TATAGTACTACCTGTGCTGAGATTCCTCTCGAAAGCTATGATTCAGAGAAACATTTG 301
QY 301 ACGGAGAGATGGAATCATACTACTTAGAGAGAAACCAAGAAGGGAAGGCATC 360

```

ABX73394  
ID ABX73394 standard; DNA; 2375 BP.  
XX  
AC ABX73394;  
XX  
DT 18-MAR-2003 (first entry)  
XX  
XX  
DE Human novel polynucleotide #222.  
XX  
XX Human; gene; ds; neural disorder; immune system disorder; renal disorder;  
KW muscular disorder; respiratory disease; reproductive disorder;  
KW gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;  
KW hyperproliferative disorder; inflammatory disease; allergic reaction;  
KW blood related disorder; cancer; immunosuppressive; antiinflammatory;  
KW cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;  
KW haemostatic; antiarteriosclerotic.  
XX  
OS Homo sapiens.  
XX  
XX US2002132753-A1.  
XX  
XX 19-SEP-2002.  
XX  
XX 17-JAN-2001; 2001US-0764864.  
XX  
XX 31-JAN-2000; 2000US-179065P.  
PR 04-FEB-2000; 2000US-180628P.  
PR 28-JUN-2000; 2000US-214686P.  
PR 07-JUL-2000; 2000US-216647P.  
PR 07-JUL-2000; 2000US-216980P.  
PR 11-JUL-2000; 2000US-217487P.  
PR 11-JUL-2000; 2000US-217496P.  
PR 14-JUL-2000; 2000US-218290P.  
PR 26-JUL-2000; 2000US-220963P.  
PR 26-JUL-2000; 2000US-220964P.  
PR 14-AUG-2000; 2000US-224518P.  
PR 14-AUG-2000; 2000US-224519P.  
PR 14-AUG-2000; 2000US-225267P.  
PR 14-AUG-2000; 2000US-225268P.  
PR 14-AUG-2000; 2000US-225270P.  
PR 14-AUG-2000; 2000US-225447P.  
PR 14-AUG-2000; 2000US-225757P.  
PR 22-AUG-2000; 2000US-225758P.  
PR 30-AUG-2000; 2000US-226868P.  
PR 01-SEP-2000; 2000US-228924P.  
PR 01-SEP-2000; 2000US-229287P.  
PR 01-SEP-2000; 2000US-229343P.  
PR 01-SEP-2000; 2000US-229345P.  
PR 05-SEP-2000; 2000US-229509P.  
PR 05-SEP-2000; 2000US-229513P.  
PR 08-SEP-2000; 2000US-231413P.  
PR 21-SEP-2000; 2000US-234223P.  
PR 21-SEP-2000; 2000US-234274P.  
PR 27-SEP-2000; 2000US-234997P.  
PR 27-SEP-2000; 2000US-235834P.  
PR 29-SEP-2000; 2000US-236327P.  
PR 29-SEP-2000; 2000US-236367P.  
PR 29-SEP-2000; 2000US-236368P.  
PR 29-SEP-2000; 2000US-236369P.  
PR 29-SEP-2000; 2000US-236370P.  
PR 02-OCT-2000; 2000US-236802P.  
PR 02-OCT-2000; 2000US-237037P.  
PR 02-OCT-2000; 2000US-237038P.  
PR 02-OCT-2000; 2000US-237039P.  
PR 02-OCT-2000; 2000US-237040P.  
PR 13-OCT-2000; 2000US-239935P.  
PR 20-OCT-2000; 2000US-240960P.  
PR 20-OCT-2000; 2000US-241785P.  
PR 20-OCT-2000; 2000US-241809P.  
PR 01-NOV-2000; 2000US-244617P.  
PR 17-NOV-2000; 2000US-249299P.  
PR 08-DEC-2000; 2000US-251856P.

PR 08-DEC-2000; 2000US-251868P.  
PR 08-DEC-2000; 2000US-251869P.  
XX  
XX (ROSE/) ROSEN C A.  
PA (RUBE/) RUBEN S M.  
PA (BARA/) BAPASH S C.  
XX  
XX Rosen CA, Ruben SM, Barash SC;  
PI  
XX  
XX WPI; 2003-147444/14.  
DR P-PSDB; ABU55134.  
XX  
XX New polypeptides and nucleic acids, useful in gene therapy for  
PT treating, inhibiting or preventing e.g. neural, immune system,  
PT muscular, respiratory, reproductive, gastrointestinal, pulmonary,  
PT cardiovascular or renal disorders -  
XX  
XX C'aim 1; SEQ ID NO 232; 402pp; English.  
XX  
XX The invention relates to human novel polypeptides and their associated  
CC polynucleotides. The polypeptides and polynucleotides are useful in gene  
CC therapy for treating, inhibiting or preventing neural disorders, immune  
CC system disorders (e.g. systemic lupus erythematosus, rheumatoid  
CC arthritis and multiple sclerosis), muscular disorders, respiratory  
CC diseases (e.g. nasal vestibulitis, nasal polyps and sinusitis),  
CC reproductive disorders, gastrointestinal disorders, pulmonary disorders,  
CC cardiovascular disorders (e.g. congenital heart defects, Ebstein's  
CC anomaly and hypoplastic left heart syndrome), renal disorders (e.g. acute  
CC kidney failure and end-stage renal disease), hyperproliferative disorders  
CC (e.g. Hodgkin's disease and leukaemia), inflammatory diseases (e.g.  
CC septic shock, bursitis and appendicitis), allergic reactions and  
CC conditions (e.g. asthma), blood related disorders (e.g. thrombosis,  
CC atherosclerosis and myocardial infarction) and cancerous diseases.  
CC Sequences ABX73173-ABX74167 represent human novel polynucleotides of the  
CC invention.  
XX  
XX  
SQ Sequence 2375 BP; 731 A; 435 C; 600 G; 609 T; 0 other;  
  
Query Match 44.7%; Score 1282; DB 25; Length 2375;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1282; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1534 AAAGGTTGAATGGTGGCCAAAGCAGACAAAGGATCACTGGATTTTATGCAAAAGGG 1643  
DB 9 AAAGGTTGAATGGTGGCCAAAGCAGACAAAGGATCACTGGATTTTATGCAAAAGGG 68  
QY 1644 TGC AAAAGGTGAAAAGGGGAATGCTGGGTTCCCTGGCTCCCTGGACCTGCTGGAGAAC 1703  
DB 69 TGC AAAAGGTGAAAAGGGGAATGCTGGGTTCCCTGGCTCCCTGGACCTGCTGGAGAAC 128  
QY 1704 AGAAGACATGGAAGGATGGATTAAATGGTAGTCCCGTTTCAAGGAGAGCAGGATC 1763  
DB 129 AGAAGACATGGAAGGATGGATTAAATGGTAGTCCCGTTTCAAGGAGAGCAGGATC 188  
QY 1764 CCTGTGTCTCCGGGCGAGGATGGAAACACGGGAGAGCTGGATCCAGGATTTCTTGG 1823  
DB 189 CCTGTGTCTCCGGGCGAGGATGGAAACACGGGAGAGCTGGATCCAGGATTTCTTGG 248  
QY 1824 AAACCGAGGATTAATGGGCCAAAGGGGAGAAATGGGCTCCAGGACAGCAAGGAAAAA 1893  
DB 249 AAACCGAGGATTAATGGGCCAAAGGGGAGAAATGGGCTCCAGGACAGCAAGGAAAAA 308  
QY 1884 AGGAGCCCCAGGATGCTGTGTTTAATGGGAGCAATGGCTCACAGGCCAGCTGGAAC 1943  
DB 309 AGGAGCCCCAGGATGCTGTGTTTAATGGGAGCAATGGCTCACAGGCCAGCTGGAAC 368  
QY 1944 ACCGGGATCTAAGGGAAGCAAAAGGTGAACCTGGGAATTCAGGGGATGCTGGGGCTTCTGG 2003  
DB 369 ACCGGGATCTAAGGGAAGCAAAAGGTGAACCTGGGAATTCAGGGGATGCTGGGGCTTCTGG 428  
QY 2004 GCTCAAGGAGAACCCAGGAGCAACGGGTTCCCGAGGAGAACCCAGGATACATGGGTTTACC 2063  
DB 429 GCTCAAGGAGAACCCAGGAGCAACGGGTTCCCGAGGAGAACCCAGGATACATGGGTTTACC 488

PR 01-DEC-2000; 2000US-0250160.  
 PR 01-DEC-2000; 2000US-0250391.  
 PR 01-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 05-DEC-2000; 2000US-0256719.  
 PR 06-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251989.  
 PR 08-DEC-2000; 2000US-0251990.  
 PR 11-DEC-2000; 2000US-0254097.  
 PR 05-JAN-2001; 2001US-0259678.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Barash SC, Ruben SM;  
 XX  
 DR WPI; 2001-488783/53.  
 DR P-PSDB; AAU16066.  
 XX  
 XX New nucleic acid molecules encoding 461 human secreted proteins for  
 PT diagnosing, preventing, treating or ameliorating medical conditions and  
 PT used as food additives or preservatives -  
 XX  
 PS Claim 1; SEQ ID No 232; 980pp; English.  
 XX  
 CC The invention relates to isolated nucleic acid molecules and their  
 CC encoded secreted proteins. The nucleic acids and proteins are used to  
 CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,  
 CC rabbits, goats, horses, cats, dogs, chickens or sheep. They  
 CC are also used in diagnosing a pathological condition or susceptibility  
 CC to a pathological condition. Antibodies to the proteins can also  
 CC be used in alleviating symptoms associated with the disorders and in  
 CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked  
 CC immunosorbant assays (ELISA). Disorders which are diagnosed or treated  
 CC include autoimmune diseases e.g. rheumatoid arthritis,  
 CC hyperproliferative disorders e.g. neoplasms of the breast or liver,  
 CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders  
 CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.  
 CC Alzheimer's disease, infections caused by bacteria, viruses and fungi  
 CC and ocular disorders e.g. corneal infection, and many other  
 CC disorders listed in the specification. The polypeptides can also  
 CC be used to aid wound healing and epithelial cell proliferation, to  
 CC prevent skin aging due to sunburn, to maintain organs before  
 CC transplantation, for supporting cell culture of primary tissues, to  
 CC regenerate tissues and in chemotaxis. The polypeptides can also be used  
 CC as a food additive or preservative to increase or decrease storage  
 CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,  
 CC minerals, cofactors and other nutritional components. The present  
 CC sequence encodes a novel secreted protein of the invention.

Query Match 44.7%; Score 1282; DB 22; Length 2375;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1282; Conservative 0; Mismatches 0; Indels 0; Gaps 3;

QY 1584 AAAGGTTGAATGGTGGCAAGAGAGACAAGGATCACTGGATTTTATGCAAAAAGG 1643  
 Db 9 AAAGGTTGAATGGTGGCAAGAGAGACAAGGATCACTGGATTTTATGCAAAAAGG 68  
 QY 1644 TCACAAGGTGAAGAGGGAATGCTGGCTTCCTCGCTCCCTGGACCTGCTGGAGACC 1703  
 Db 69 TCACAAGGTGAAGAGGGAATGCTGGCTTCCTCGCTCCCTGGACCTGCTGGAGACC 128  
 QY 1704 AGGAAGACATGAAAGGATGGATTAATGGTGTAGTCCCGTTTCAAGGGAGAGCAGGATC 1763  
 Db 129 AGGAAGACATGAAAGGATGGATTAATGGTGTAGTCCCGTTTCAAGGGAGAGCAGGATC 188  
 QY 1764 CCTGTGTCTCCGGGACAGGATGGAACACGGGAGAGCTTGAATCCCAAGGATTTCTGTG 1823  
 Db 189 CCTGTGTCTCCGGGACAGGATGGAACACGGGAGAGCTTGAATCCCAAGGATTTCTGTG 248  
 QY 1824 AAACCGAGGATTAATGGGCCAAAAGGAGAAATGGGGCTCCAGACAGCAAGGAAAAA 1883

Db 249 AAACCGAGGATTAATGGGCCAAAAGGAGAAATGGGGCTCCAGACAGCAAGGAAAAA 308  
 QY 1884 AGGAGCCCCAGGATGCTGGTTTAAATGGGAAGCAATGGCTCACAGGCCAGCCTGGAC 1943  
 Db 309 AGGAGCCCCAGGATGCTGGTTTAAATGGGAAGCAATGGCTCACAGGCCAGCCTGGAC 368  
 QY 1944 ACCGGGATCTAAGGGAAGCAAAAGGTGAACCTGGAATTCAAAGGATGCTGGGGCTTCTGG 2003  
 Db 369 ACCGGGATCTAAGGGAAGCAAAAGGTGAACCTGGAATTCAAAGGATGCTGGGGCTTCTGG 428  
 QY 2004 GCTCAAGGAGAAACCCAGGAGCAACCGGTTCCCAAGGAGAACCCAGGATACATGGTTTACC 2063  
 Db 429 GCTCAAGGAGAAACCCAGGAGCAACCGGTTCCCAAGGAGAACCCAGGATACATGGTTTACC 488  
 QY 2064 CGGGATTCAGGAAAAAAGGGGGAACAAAGGAATCAAGGTGAAGGAATTCAGGGCTCATCG 2123  
 Db 489 CGGGATTCAGGAAAAAAGGGGGAACAAAGGAATCAAGGTGAAGGAATTCAGGGCTCATCG 548  
 QY 2124 AAAGGAGAAAAATGGAAGACAGGGGAATTCAGGGCAACAGGGAATTCAGGGCTCATCG 2183  
 Db 549 AAAGGAGAAAAATGGAAGACAGGGGAATTCAGGGCAACAGGGAATTCAGGGCTCATCG 608  
 QY 2184 TGCAAAAGSAGAGAGAGGTGAAGGAGAACCTGGTGTCCAGAGTGCATTTGATTCAAA 2243  
 Db 609 TGCAAAAGSAGAGAGAGGTGAAGGAGAACCTGGTGTCCAGAGTGCATTTGATTCAAA 668  
 QY 2244 AGGAGAACTCGGGGTGGATGGCTTGAATGGGGCCAGAGTCTTAAGGGCAACCTGGGA 2303  
 Db 669 AGGAGAACTCGGGGTGGATGGCTTGAATGGGGCCAGAGTCTTAAGGGCAACCTGGGA 728  
 QY 2304 TCCAGTCTCTCAGGGACCCCGGAGTTTGGATGGGAAGCCCGGAAGAGAGTTTTCAGAAC 2363  
 Db 729 TCCAGTCTCTCAGGGACCCCGGAGTTTGGATGGGAAGCCCGGAAGAGAGTTTTCAGAAC 788  
 QY 2364 ATTATTCAGCAAGTTTGGACAGATGTATTAAGCCAGCTACAGTCTTACTTCAGAG 2423  
 Db 789 ATTATTCAGCAAGTTTGGACAGATGTATTAAGCCAGCTACAGTCTTACTTCAGAG 848  
 QY 2424 TGGAGAAATAGAAATTTGTGATCATTCGCTGTCCCAACATGCTCCCGGGTATTCCTGG 2483  
 Db 849 TGGAGAAATAGAAATTTGTGATCATTCGCTGTCCCAACATGCTCCCGGGTATTCCTGG 908  
 QY 2484 GCCACCTGTTCGATAGGCCCCAGAGGTTCCAGAGGATTAACCTGGTTTCCAGGAAGAGA 2543  
 Db 909 GCCACCTGTTCGATAGGCCCCAGAGGTTCCAGAGGATTAACCTGGTTTCCAGGAAGAGA 968  
 QY 2544 TGGTGTTCCTGGATTTAGTGGGTGTCCTGGAGCTCCAGGTGTCAGAGGATTAAGAGGCT 2603  
 Db 969 TGGTGTTCCTGGATTTAGTGGGTGTCCTGGAGCTCCAGGTGTCAGAGGATTAAGAGGCT 1028  
 QY 2604 ACCAGGAAGAAATGGGGAAAAAGGGAGCCAAAGGTTTGGGTATCTCGGAGAACAGGTC 2663  
 Db 1029 ACCAGGAAGAAATGGGGAAAAAGGGAGCCAAAGGTTTGGGTATCTCGGAGAACAGGTC 1088  
 QY 2664 TCTGTGTCCCGAGGTCCAGAGGCCCTCTCGAATTAAGCAAGAGAGGTCTCTCCAGGAGA 2723  
 Db 1089 TCTGTGTCCCGAGGTCCAGAGGCCCTCTCGAATTAAGCAAGAGAGGTCTCTCCAGGAGA 1148  
 QY 2724 CCCAGGTCTCCTGGCAAGATGGAGACCATGGAACCTGGAATCCAGGGCAACCCAGG 2783  
 Db 1149 CCCAGGTCTCCTGGCAAGATGGAGACCATGGAACCTGGAATCCAGGGCAACCCAGG 1208  
 QY 2784 CCCCCAGGCATCTGGACCCCATCACTATGTTTATGTGTAATTGCGCAGAGAGATCCGTT 2843  
 Db 1209 CCCCCAGGCATCTGGACCCCATCACTATGTTTATGTGTAATTGCGCAGAGATCCGTT 1268  
 QY 2844 CAGAAAAGGACCAAACTATTAG 2865  
 Db 1269 CAGAAAAGGACCAAACTATTAG 1290

RESULT 4

XX 07-NOV-2001 (first entry)  
XX Human cDNA encoding a novel secreted protein, Seq ID 232.  
XX  
XX Human; immunosuppressive; antiarthritic; ss; anti-rheumatic;  
KW cytosolic; cardiac; vasotropic; cerebroprotective; nootropic;  
KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;  
KW vulnary; secreted protein; rheumatoid arthritis;  
KW hyperproliferative disorder; cardiovascular disorder; cardiac arrest;  
KW cerebrovascular disorder; cerebral ischaemia; angiogenesis;  
KW nervous system disorder; Alzheimer's disease; infection; ocular disorder;  
KW corneal infection; wound healing; epithelial cell proliferation;  
KW skin ageing; food additive; preservative; antiproliferative.  
OS Homo sapiens.  
XX  
XX WO200155322-A2.  
XX  
XX 02-AUG-2001.  
XX  
XX 17-JAN-2001; 2001WO-US01341.  
XX  
XX 31-JAN-2000; 2000US-0179065.  
XX 04-FEB-2000; 2000US-0180628.  
XX 24-FEB-2000; 2000US-0184664.  
XX 02-MAR-2000; 2000US-0186350.  
XX 16-MAR-2000; 2000US-0189874.  
XX 17-MAR-2000; 2000US-0190076.  
XX 18-APR-2000; 2000US-0198123.  
XX 19-MAY-2000; 2000US-0205515.  
XX 07-JUN-2000; 2000US-0209467.  
XX 28-JUN-2000; 2000US-0214886.  
XX 30-JUN-2000; 2000US-0215135.  
XX 07-JUL-2000; 2000US-0216647.  
XX 07-JUL-2000; 2000US-0216880.  
XX 11-JUL-2000; 2000US-0217487.  
XX 11-JUL-2000; 2000US-0217496.  
XX 14-JUL-2000; 2000US-0218290.  
XX 26-JUL-2000; 2000US-0220963.  
XX 26-JUL-2000; 2000US-0220964.  
XX 14-AUG-2000; 2000US-0224518.  
XX 14-AUG-2000; 2000US-0224519.  
XX 14-AUG-2000; 2000US-0225213.  
XX 14-AUG-2000; 2000US-0225214.  
XX 14-AUG-2000; 2000US-0225266.  
XX 14-AUG-2000; 2000US-0225267.  
XX 14-AUG-2000; 2000US-0225268.  
XX 14-AUG-2000; 2000US-0225270.  
XX 14-AUG-2000; 2000US-0225447.  
XX 14-AUG-2000; 2000US-0225757.  
XX 14-AUG-2000; 2000US-0225758.  
XX 14-AUG-2000; 2000US-0225759.  
XX 18-AUG-2000; 2000US-0226279.  
XX 22-AUG-2000; 2000US-0226681.  
XX 22-AUG-2000; 2000US-0226868.  
XX 22-AUG-2000; 2000US-0227182.  
XX 23-AUG-2000; 2000US-0227009.  
XX 30-AUG-2000; 2000US-0228924.  
XX 01-SEP-2000; 2000US-0229287.  
XX 01-SEP-2000; 2000US-0229343.  
XX 01-SEP-2000; 2000US-0229344.  
XX 05-SEP-2000; 2000US-0229345.  
XX 05-SEP-2000; 2000US-0229509.  
XX 06-SEP-2000; 2000US-0229513.  
XX 06-SEP-2000; 2000US-0230437.  
XX 06-SEP-2000; 2000US-0230438.  
XX 08-SEP-2000; 2000US-0231242.  
XX 08-SEP-2000; 2000US-0231243.  
XX 08-SEP-2000; 2000US-0231244.  
XX 08-SEP-2000; 2000US-0231413.  
XX 08-SEP-2000; 2000US-0231414.  
XX 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 23-OCT-2000; 2000US-0241787.  
PR 23-OCT-2000; 2000US-0241808.  
PR 23-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 31-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.

PR 03-FEB-2000; 2000US-0496914.  
 PR 21-APR-2000; 2000US-0560875.  
 PR 20-JUN-2000; 2000US-0598075.  
 PR 19-JUL-2000; 2000US-0620325.  
 PR 01-SEP-2000; 2000US-0654936.  
 PR 15-SEP-2000; 2000US-0663561.  
 PR 20-OCT-2000; 2000US-0693325.  
 PR 30-NOV-2000; 2000US-0728422.  
 XX (HYSE-) HYSEQ INC.  
 PA  
 XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;  
 PI Zhao QA, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
 PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;  
 XX  
 DR WPI; 2001-476283/51.  
 DR P-PSDB; AAM79110.  
 XX  
 XX Nucleic acids encoding polypeptides with cytokine-like activities,  
 PT useful in diagnosis and gene therapy .  
 XX  
 PS Claim 1; Page 2635-2637; 6221pp; English.  
 XX  
 CC The invention relates to polynucleotides (AAK51456-AAK53435) and the  
 CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activin/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation.  
 CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666  
 CC (AAM80020) are omitted as the relevant pages from the sequence listing  
 CC were missing at the time of publication.  
 XX  
 XX Sequence 1826 BP; 578 A; 337 C; 418 G; 493 T; C other;  
 SQ  
 Query Match 49.8%; Score 1426; DB 22; Length 1826;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1426; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ATGGCTCACTATATACATTTCTCTGATGGTTTGGTGTCTTCTTCTGAGATTTCTGTG 60  
 DB 2 ATGGCTCACTATATACATTTCTCTGATGGTTTGGTGTCTTCTTCTGAGATTTCTGTG 61  
 QY 61 TTAGCTCAAGATGGGAGTAAGATCAAGTTGTCTGCTCTCCGACAGATTAGTTTTC 120  
 DB 62 TTAGCTCAAGATGGGAGTAAGATCAAGTTGTCTGCTCTCCGACAGATTAGTTTTC 121  
 QY 121 ATCTTAGATGCTCTTATAGTGTGGCCGAGAAATCTTTGAAATAGTGAAGAGTGGCTT 180  
 DB 122 ATCTTAGATGCTCTTATAGTGTGGCCGAGAAATCTTTGAAATAGTGAAGAGTGGCTT 181  
 QY 181 GTCAATATCAAAAACCTTTGATGATGCGGCGGAGTTTATTTCAAGTTGGAGTGTCAA 240  
 DB 182 GTCAATATCAAAAACCTTTGATGATGCGGCGGAGTTTATTTCAAGTTGGAGTGTCAA 241  
 QY 241 TATAGTACTACCTCTGCTGGAGATTCCTCTCGGAAGCTATGATTCAGGAGAACATTTG 300  
 DB 242 TATAGTACTACCTCTGCTGGAGATTCCTCTCGGAAGCTATGATTCAGGAGAACATTTG 301  
 QY 301 ACCGACAGTGGATCCATCTCTTCTAGGAGGAACACAAAGACAGGAGGCGCATC 360  
 DB 302 ACCGACAGTGGATCCATCTCTTCTAGGAGGAACACAAAGACAGGAGGCGCATC 361  
 QY 361 CAGTTTGGCTCGATTTACCTTTTGGCAAGTCTCTCAAGTTTCTGACTAAGATAGCAGTG 420  
 DB 362 CAGTTTGGCTCGATTTACCTTTTGGCAAGTCTCTCAAGTTTCTGACTAAGATAGCAGTG 421  
 QY 421 GTACTTAGGATGCAATCCCAAGATGACGTCAAGGATGCACTCAAGCAGCAGAGAT 480

DB 422 GTACTTAGGATGCAATCCCAAGATGACGTCAAGGATGCAAGTCAAGCAGCAGAGAT 481  
 QY 481 AGTAAGATAACATTATTGCTATTGCTGTTGGTTTCTGAGAAACAGAGATGCCGAACTTAGA 540  
 DB 482 AGTAAGATAACATTATTGCTATTGCTGTTGGTTTCTGAGAAACAGAGATGCCGAACTTAGA 541  
 QY 541 GCTATTGCCAAGAGGCTTGGTCTACTTATGTTGTTTATGTTGGAAGACTATATTCGAATA 600  
 DB 542 GCTATTGCCAAGAGGCTTGGTCTACTTATGTTGTTTATGTTGGAAGACTATATTCGAATA 601  
 QY 601 TCCAAATTAAGGAAAGTGAAGAGCAGAAATCTTGTGAAGAACTCTGCTGTCCCAACAGA 660  
 DB 602 TCCAAATTAAGGAAAGTGAAGAGCAGAACTTGTGAAGAACTCTGCTGTCCCAACAGA 661  
 QY 661 ATTCCAGTGGCAGCTCGTATGAAGGGGATTTGATATTTCTTTTAGGTTTAGATGTAAT 720  
 DB 662 ATTCCAGTGGCAGCTCGTATGAAGGGGATTTGATATTTCTTTTAGGTTTAGATGTAAT 721  
 QY 721 AAAAAGCTTAAGAAAGAAATACAGCTTCCACCAAAAAGATTAAGAGGATATGAAGTAACA 780  
 DB 722 AAAAAGCTTAAGAAAGAAATACAGCTTCCACCAAAAAGATTAAGAGGATATGAAGTAACA 781  
 QY 781 TCAAAAGTGTATTTATCAGAACTCAGAACTCAAGCAATGTTTTCCCAAGAGGCTTCTCTCATCA 840  
 DB 782 TCAAAAGTGTATTTATCAGAACTCAGAACTCAGAACTGTTTTCCCAAGAGGCTTCTCTCATCA 841  
 QY 841 TATGCTATTTGCTCTACTCAGAGATTTAAGTCAGAAATTTGGGATTTATGGAATA 900  
 DB 842 TATGCTATTTGCTCTACTCAGAGATTTAAGTCAGAAATTTGGGATTTATGGAATA 901  
 QY 901 TTAATATTGATGGAAGGCAACAAATAGCAGTACCTTAAATGCTGTGACCAAAATCTTA 960  
 DB 902 TTAATATTGATGGAAGGCAACAAATAGCAGTACCTTAAATGCTGTGACCAAAATCTTA 961  
 QY 961 TTAATATTGATGGAAGGCAACAAATAGCAGTACCTTAAATGCTGTGACCAAAATCTTA 962  
 DB 962 TTAATATTGATGGAAGGCAACAAATAGCAGTACCTTAAATGCTGTGACCAAAATCTTA 1020  
 QY 1021 GTTAAAGACGTTGTTGATGAAGGCTGGCACCACAAATTCGTTCTTTAGTAAACAGAAACAGAT 1080  
 DB 1022 GTTAAAGACGTTGTTGATGAAGGCTGGCACCACAAATTCGTTCTTTAGTAAACAGAAACAGAT 1081  
 QY 1081 GTGACTTTGTATATTGATGACCAAAATTTGAAACAGACCTTACATCCAGTTTATAGG 1140  
 DB 1082 GTGACTTTGTATATTGATGACCAAAATTTGAAACAGACCTTACATCCAGTTTATAGG 1141  
 QY 1141 ATCTTGATCAATGGGCAACCCAAATTTGGAATAATTTCTGGAAAGAGAAACTGTTCAG 1200  
 DB 1142 ATCTTGATCAATGGGCAACCCAAATTTGGAATAATTTCTGGAAAGAGAAACTGTTCAG 1201  
 QY 1201 TTTGATGTCACAAAGTTGCGAATCTACTGTGACCCAGAAACAGAAACCGGGAGACAGCA 1260  
 DB 1202 TTTGATGTCACAAAGTTGCGAATCTACTGTGACCCAGAAACAGAAACCGGGAGACAGCA 1261  
 QY 1261 TGTGAGATTCCTCGATTTGCTTATGCTCCAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGT 1320  
 DB 1262 TGTGAGATTCCTCGATTTGCTTATGCTCCAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGT 1321  
 QY 1321 TGTATTGTCTCCGGGAAACCCAGGACTTCAAGGCCCCCAAGAGTGTAGTGTAGTGTAGTGTAGT 1380  
 DB 1322 TGTATTGTCTCCGGGAAACCCAGGACTTCAAGGCCCCCAAGAGTGTAGTGTAGTGTAGTGTAGT 1381  
 QY 1381 GGGAAACCTTGCTACCTCGACAACTGTGATGATGATGATGATGATGATGATGATGATGATGAT 1426  
 DB 1382 GGGAAACCTTGCTACCTCGACAACTGTGATGATGATGATGATGATGATGATGATGATGATGAT 1427

RESULT 3  
 AAK526053 standard; cDNA; 2375 bp.  
 XX  
 AC AAK526053;

1223 GTTAAGACGTTGTTGATGAAGGCTGGCACCAATTCGTCCTCTAGTAACAGAACAGAT 1282  
1081 GTGACTTTGTAATATGATGACCAACAAATTTGAACAAGCCCTTACATCCAGTTTATGGG 1340  
1283 GTGACTTTGTAATATGATGACCAACAAATTTGAACAAGCCCTTACATCCAGTTTATGGG 1342  
1141 ATCTTGATCAATGGGCACACCCAAATTTGAAATATCTGTGAAAGAGAACTGTTTCAG 1200  
1343 ATCTTGATCAATGGGCACACCCAAATTTGAAATATCTGTGAAAGAGAACTGTTTCAG 1402  
1201 TTTTGATGTCCTCAAAAGTTGGAATCTACTGTGACCCAGAACAGAACACCGGAGACGCA 1260  
1403 TTTTGATGTCCTCAAAAGTTGGAATCTACTGTGACCCAGAACAGAACACCGGAGACGCA 1462  
1261 TGTGAGATTCCTGGATTT-----TGCCCTTAATGCTCCAGTGATGTAGGTTCAACT 1311  
1463 TGTGAGATTCCTGGATTTAATGAGAGTGCCCTTAATGCTCCAGTGATGTAGGTTCAACT 1522  
1312 CCAGCTCCCTGTATTTGCTCCCGGAAAAACAGGACTTCAAGGCCCCCAAGAGTGACCCCT 1371  
1523 CCAGCTCCCTGTATTTGCTCCCGGAAAAACAGGACTTCAAGGCCCCCAAGAGTGACCCCT 1582  
1372 GGAATGCTGGGAACCTGGCTACCTTGACAACTGCTCAAGTGAAGTGAAGCTGGATAT 1431  
1583 GGAATGCTGGGAACCTGGCTACCTTGACAACTGCTCAAGTGAAGTGAAGCTGGATAT 1642  
1432 CAGGAAATGACAGGACACCAAGTGCTCCAGGATCTCCAGGAATACAAAGAGCTCGAGGA 1491  
1643 CAGGAAATGACAGGACACCAAGTGCTCCAGGATCTCCAGGAATACAAAGAGCTCGAGGA 1702  
1492 CTACAGGTTACAAAGAGAACCAAGGAGATGTGTGACAAAGGTTGATCGTGACTTCCT 1551  
1703 CTACAGGTTACAAAGAGAACCAAGGAGATGTGTGACAAAGGTTGATCGTGACTTCCT 1762  
1552 GGTGTTCTCGGCTTCATGGATGACAGATCAAGAGGTGAATGGGTGCCAAAGGAGAC 1611  
1763 GGTGTTCTCGGCTTCATGGATGACAGATCAAGAGGTGAATGGGTGCCAAAGGAGAC 1822  
1612 AAAGATACCTCGATTTATGCAAAAGAGGTGCAAAGGTGAAAGGGGAATGCTGGC 1671  
1823 AAAGATACCTCGATTTATGCAAAAGAGGTGCAAAGGTGAAAGGGGAATGCTGGC 1882  
1672 TCCCTGGCTCCCTGGACCTCTCGAGAACAGAGACATGAAAGGATGGAATTAATG 1731  
1883 TCCCTGGCTCCCTGGACCTCTCGAGAACAGAGACATGAAAGGATGGAATTAATG 1842  
1732 GGTAGTCCCGGTTTCAAGGAGAGCAGGATCCCTGTGTCTCCGGGACAGGATGGAACA 1791  
1943 GGTAGTCCCGGTTTCAAGGAGAGCAGGATCCCTGTGTCTCCGGGACAGGATGGAACA 2002  
1792 CGGGAGAGCCTGGAATCCAGGATTTCTGGAACCCAGGATTAATGGGCCAAAGGGA 1851  
2003 CGGGAGAGCCTGGAATCCAGGATTTCTGGAACCCAGGATTAATGGGCCAAAGGGA 2062  
1852 GAAAATGGGCTCCAGGACAGCAAGAAAAAGAGCCCGGATGCTGGTTTAATG 1911  
2063 GAAAATGGGCTCCAGGACAGCAAGAAAAAGAGCCCGGATGCTGGTTTAATG 2122  
1912 GGAAGCAATGGCTACACCGGACCTGGAACACCGGATCTAAGGGAAGCAAGGTGAA 1971  
2123 GGAAGCAATGGCTACACCGGACCTGGAACACCGGATCTAAGGGAAGCAAGGTGAA 2182  
1972 CTGGAATTCAGGATGCTCGGGCTTCAGGGCTCAAGGGAGAACAGGAGAACCGGT 2031  
2183 CTGGAATTCAGGATGCTCGGGCTTCAGGGCTCAAGGGAGAACAGGAGAACCGGT 2242  
2032 TCCCCAGGAGAACAGGATACATGGGTTTACCGGGATTCAGGAAAAAAGGGGACAAA 2091  
2243 TCCCCAGGAGAACAGGATACATGGGTTTACCGGGATTCAGGAAAAAAGGGGACAAA 2302  
2092 GGAATTCAGGATGGAAGGATTCAGGCTCAAGGGAGAAATGGAAGACAGGGAATT 2151  
2303 GGAATTCAGGATGGAAGGATTCAGGCTCAAGGGAGAAATGGAAGACAGGGAATT 2362

2152 CCAGGGCAACAGGGAATTCAGAGGCATCATGTCGAAAAAGAGAGAGAGGTGAAAAGGGA 2211  
2363 CCAGGGCAACAGGGAATTCAGAGGCATCATGTCGAAAAAGAGAGAGAGGTGAAAAGGGA 2422  
2212 GAACCTGCTGTCGAGGTGCCATTTGGATCAAAAGAGAACTTGGGTGGATGGCTTCATG 2271  
2423 GAACCTGCTGTCGAGGTGCCATTTGGATCAAAAGAGAACTTGGGTGGATGGCTTCATG 2482  
2272 GGGCCGACAGGTGCTTAAAGGCAACCTGGGATCCAGTCTCTCAGGACCCCAAGGTTG 2331  
2483 GGGCCGACAGGTGCTTAAAGGCAACCTGGGATCCAGTCTCTCAGGACCCCAAGGTTG 2542  
2332 GAAGGAGAGCCCGGAAGAGAGTTTCAGACAAATTTGCAAGTTTCACAGATGTA 2391  
2543 GAAGGAGAGCCCGGAAGAGAGTTTCAGACAAATTTGCAAGTTTCACAGATGTA 2602  
2392 ATAAGAGCCCAAGTACCAAGTCTTACTTCAGAGTGAAGAAATTAGAAATTGTGATCATTC 2451  
2603 ATAAGAGCCCAAGTACCAAGTCTTACTTCAGAGTGAAGAAATTAGAAATTGTGATCATTC 2662  
2452 CTGTCGCAACATGCTCCCGGTATTCCTGGGCACTGCTGGTCCGATAGGCCCAAGGTTG 2511  
2663 CTGTCGCAACATGCTCCCGGTATTCCTGGGCACTGCTGGTCCGATAGGCCCAAGGTTG 2722  
2512 CCCAGAGATTACCTGTTTCCAGGAAAGAGATGTTCTCTGATTTAGTGGGTGTCCT 2571  
2723 CCCAGAGATTACCTGTTTCCAGGAAAGAGATGTTCTCTGATTTAGTGGGTGTCCT 2782  
2572 GGAATCCAGTGTTCAGAGATTTAAAGGCTTACAGGAAGAAATGGGAAAAAGGAGC 2631  
2783 GGAATCCAGTGTTCAGAGATTTAAAGGCTTACAGGAAGAAATGGGAAAAAGGAGC 2842  
2632 CAAAGGTTTGGGTATCTCGGAGAACAGGTCCTCTGTCCTCCAGGTCCAGAGGGCCCT 2691  
2843 CAAAGGTTTGGGTATCTCGGAGAACAGGTCCTCTGTCCTCCAGGTCCAGAGGGCCCT 2902  
2692 CCGTAATTAAGCAAGAGGTCTCCAGGAGACCCAGGTCTCTCTGCAAGATGGAGAC 2751  
2903 CCGTAATTAAGCAAGAGGTCTCCAGGAGACCCAGGTCTCTCTGCAAGATGGAGAC 2962  
2752 CATGAAACCTGGAATCCAGGGCAACAGGCTCCAGGACCCCAAGCACTATTAG 2811  
2963 CATGAAACCTGGAATCCAGGGCAACAGGCTCCAGGACCCCAAGCACTATTAG 3022  
2812 TGCTTTAGTGAATTTGCGAAGAGATCCGTTCAAGAAAAAGGACCAAACTATTAG 2865  
3023 TGCTTTAGTGAATTTGCGAAGAGATCCGTTCAAGAAAAAGGACCAAACTATTAG 3076

## RESULT 2

AAK52243  
ID AAK52243 standard: cDNA; 1826 bp.

XX AAK52243;

XX AC AAK52243;

XX AC AAK52243;

DT 06-NOV-2001 (first entry)

XX Human polynucleotide SEQ ID NO 788.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
XX tissue growth factor; immunomodulatory; cancer; leukaemia;  
XX nervous system disorder; arthritis; inflammation; ss.

CS Homo sapiens.

XX WO200157190-A2.

XX 09-AUG-2001.

XX 05-FEB-2001; 2001WO-US04098.

XX

19 0.7 2787 23 AAS77413 DNA encoding novel  
 19 0.7 2787 23 AAS77593 DNA encoding novel  
 19 0.7 2787 23 AAS82312 DNA encoding novel  
 19 0.7 2845 18 AAT93093 Human GCFl cDNA.  
 19 0.7 3188 23 ABU19112 Drosophila melanog  
 19 0.7 3423 20 AAX40183 Lung cancer associ  
 19 0.7 3523 18 AAT93091 Human transcriptio  
 19 0.7 3600 22 AAK85958 Human immune/haem  
 19 0.7 3608 25 ABT18028 Aspergillus fumiga  
 19 0.7 3609 25 ABT19842 Aspergillus fumiga  
 19 0.7 6741 22 AAK85961 Human immune/haem  
 19 0.7 8766 24 AAK85961 Streptococcus agal  
 19 0.7 10437 22 AAK85961 Genomic sequence #  
 19 0.7 12620 22 ABA16691 Human nervous syst  
 19 0.7 48000 22 AAF27996 Human calcium sens  
 19 0.7 66685 22 AAS07390 Human genomic DNA  
 19 0.7 66686 24 ABS73149 Human CLASP-5 geno  
 19 0.7 160755 23 AAH88704 Human DNA sequence  
 19 0.7 910715 20 AAX20248 Borrelia burgdorfe

ALIGNMENTS

RESULT 1  
 ABX71234  
 ID ABX71234 standard; cDNA; 4160 BP.

15-APR-2003 (first entry)  
 Cell structure and mobility-associated cDNA from clone DKFZphfbr2\_2b5.

Human; gene; gene therapy; vaccine; disease treatment; detection; ss.  
 Homo sapiens.

WO200112659-A2.  
 22-FEB-2001.

18-AUG-2000; 2000WO-1B01496.  
 18-AUG-1999; 99US-0149499.  
 28-SEP-1999; 99US-0156503.

(GEHU-) GERMAN HUMAN GENOME PROJECT.  
 Wiemann S;

WPI: 2001-327840/34.  
 P-PSDB; ABUS2683.

Nucleic acids having the sequences of clones isolated from libraries of  
 different human tissues, useful in recombinant DNA methodologies -

Claim 1; Page 201-202; 1095pp; English.

This invention describes novel polynucleotides and polypeptides isolated  
 from human cDNA libraries which can be used for gene therapy or in  
 vaccines. The polynucleotides of the invention and antibodies encoded by  
 them may be used in the prevention, diagnosis and treatment of diseases  
 associated with inappropriate polypeptide expression. The products of the  
 invention may also be used to identify modulators of expression and  
 activity and to down regulate expression and activity. The antibodies of  
 the invention may also be used as diagnostic agents for detecting the  
 presence of polypeptides in samples. This sequence encodes a polypeptide  
 described in the disclosure of the invention.

Sequence 4160 BP; 1271 A; 795 C; 1026 G; 1068 T; 0 other;

Query Match 71.9%; Score 2061; DB 23; Length 4160;

Best Local Similarity 99.5%; Pred No. 0;  
 Matches 2861; Conservative 0; Mismatches 4; Indels 9; Gaps 1;  
 0  
 203 ATGGCTCAGTATATTAACATTTCTGTCATGTTTGGTGGTCTCTTCTCAGAAATCTCTG 60  
 203 ATGGCTCAGTATATTAACATTTCTGTCATGTTTGGTGGTCTCTTCTCAGAAATCTCTG 262  
 61 TTAGCTGAAGATGGGAGTAAGATCAAGTCTGCTACTGCTCCGACAGATTTAGTTTTC 120  
 263 TTAGCTGAAGATGGGAGTAAGATCAAGTCTGCTACTGCTCCGACAGATTTAGTTTTC 322  
 121 ATCTTAGATGGCTTTATAGTGTGGCCAGAAAACCTTTGAAATAGTCAAAAAGTGGCT 180  
 323 ATCTTAGATGGCTTTATAGTGTGGCCAGAAAACCTTTGAAATAGTCAAAAAGTGGCT 382  
 181 GTCATATACAAAAAANCTTGCATAGAGCGGAGTATATCAAGTGGAGTGGTTCAA 240  
 383 GTCATATACAAAAAANCTTGCATAGAGCGGAGTATATCAAGTGGAGTGGTTCAA 442  
 241 TATAGTGACTACCTGTGCTGGAGATTCCTCTCGGAAGCTATGATTCAGGAGAAATTTG 300  
 443 TATAGTGACTACCTGTGCTGGAGATTCCTCTCGGAAGCTATGATTCAGGAGAAATTTG 502  
 301 ACGGCAGCAGTGGATCCATCTACTTACAGGAGAAACACAAAGACAGGAGGCCATC 360  
 503 ACGGCAGCAGTGGATCCATCTACTTACAGGAGAAACACAAAGACAGGAGGCCATC 562  
 361 CAGTTTGGCTGATTTACCTTTTTCGCAAGTCTCTCAGAGTCTTCTGACTAAGATAGCAGT 420  
 563 CAGTTTGGCTGATTTACCTTTTTCGCAAGTCTCTCAGAGTCTTCTGACTAAGATAGCAGT 622  
 421 GTACTTACCGATGGGAAATCCCAAGATCGCTCAGGATGCGCTCAAGCAGCAAGAGAT 480  
 623 GTACTTACCGATGGGAAATCCCAAGATCGCTCAGGATGCGCTCAAGCAGCAAGAGAT 682  
 481 AGTAAGATTAACATTAATTCGCTATTTGGTGTTCGACAGAAACAGAAAGATGCGCAACTAGA 540  
 683 AGTAAGATTAACATTAATTCGCTATTTGGTGTTCGACAGAAACAGAAAGATGCGCAACTAGA 742  
 541 GCTATTGCCAACAGCCTTCGCTACTTATGTTTATGTTGTTTATGTTGAAAGACTATATTGCAATA 600  
 743 GCTATTGCCAACAGCCTTCGCTACTTATGTTTATGTTGTTTATGTTGAAAGACTATATTGCAATA 802  
 601 TCCAAATTAAGGGAGTGTATGAAGCAGAAACCTTGTGAGAGATCTGTCTGCCACACAGA 660  
 803 TCCAAATTAAGGGAGTGTATGAAGCAGAAACCTTGTGAGAGATCTGTCTGCCACACAGA 862  
 661 ATTCCAGTGGCAGCTCGTGATCAAAAGGGGATTTGATATTTCTTTAGGTTTAGATGTAAT 720  
 863 ATTCCAGTGGCAGCTCGTGATCAAAAGGGGATTTGATATTTCTTTGGGTTTAGATGTAAT 922  
 721 AAAAAGGTTAAGAAAGATACAGCTTTCACAAAAGATTAAGAGGATATGAAGTAACA 780  
 923 AAAAAGGTTAAGAAAGATACAGCTTTCACAAAAGATTAAGAGGATATGAAGTAACA 982  
 781 TCAAAAGTTGATTTATCAGAACTCAAGCAATGTTTCCCGAAGGCTTCTCTCCATCA 840  
 983 TCAAAAGTTGATTTATCAGAACTCAAGCAATGTTTCCCGAAGGCTTCTCTCCATCA 1042  
 841 TATGTTATTTGCTCTACTCAAGATTTAAAGTCAAGAAAATTTGGGATTTATGAGAAATA 900  
 1043 TATGTTATTTGCTCTACTCAAGATTTAAAGTCAAGAAAATTTGGGATTTATGAGAAATA 1102  
 901 TTAACCTATTGATGGAGGCCCAAAATACAGTTACCTTTAAATGGTGGGACAAATCTTA 960  
 1103 TTAACCTATTGATGGAGGCCCAAAATACAGTTACCTTTAAATGGTGGGACAAATCTTA 1162  
 961 TTAATTAACCAACCCGCTAATTAATGCTCACAAGTGGTTACCTTTGCTTAACCCCTCAA 1020  
 1163 TTAATTAACCAACCCGCTAATTAATGCTCACAAGTGGTTACCTTTGCTTAACCCCTCAA 1222  
 1021 GTTAAGACGCTGTTTGTATGAAGGCTGGCAACAAATTCGCTCTCTAGTAACAGAAAGAT 1080

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2003, 19:26:03 ; Search time 715 Seconds  
(without alignments)  
10816.629 Million cell updates/sec

Title: US-09-996-611B-5  
Perfect score: 2865  
Sequence: 1 atgggtcattattacatt.....gaaaggacaaactattag 2865

Scoring table: OLIGO\_NUC  
Gapop\_60.0, Capext 60.0

Searched: 2552756 seqs, 1349719017 residues

Word size: 0

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database: N Geneseq 19Jun03.\*

1:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*
25:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2061	71.9	4160	23	ABX71234 Cell structure and
2	1426	49.8	1826	22	AAK52243 Human polynucleoti
3	1282	44.7	2375	22	AAS26053 Human CDNA encodin
4	1282	44.7	2375	25	ABX73394 Human novel polynu
5	1254	43.8	1578	22	AAK53227 Human polynucleoti
6	1254	43.8	1578	22	AAK53249 Human polynucleoti
7	1203	42.0	1691	22	AAK52265 Human polynucleoti
8	1062	37.1	2230	22	AAK52045 Human polynucleoti

9	1062	37.1	2230	22	AA158402	Human polynucleoti
10	974	34.0	1184	24	ABQ72640	Human MDDT encodin
11	966	33.7	2209	22	AAK53029	Human polynucleoti
12	966	33.7	2209	22	AA160188	Human polynucleoti
13	923	32.2	1169	24	ABQ72511	Human MDDT encodin
14	749	26.1	1115	22	AAS31132	Human diagnostic a
15	382	13.3	1414	24	ABX71703	Human dithp polynu
16	252	8.8	447	20	AAV86274	EST clone AJ54. H
17	166	5.8	221	22	AA149752	Probe #18438 used
18	131	4.6	246	21	AAK29851	Human secreted pro
19	52	2.9	181	22	ABA69546	Human foetal liver
20	63	2.2	1101	24	ABK71706	Human dithp polynu
21	28	1.0	466	22	AA336639	Probe #5325 used t
22	22	0.8	2154	22	AAK81046	Human immune/haema
23	20	0.7	393	21	AAK43283	Arabidopsis thalia
24	20	0.7	393	24	AB214620	Arabidopsis thalia
25	20	0.7	479	22	ABA57633	Human foetal liver
26	20	0.7	479	22	AAK05687	Human brain expres
27	20	0.7	479	22	AAK31302	Human bone marrow
28	20	0.7	479	22	AA137199	Probe #5885 used t
29	20	0.7	479	23	ABS30983	Human liver single
30	20	0.7	479	24	ABS06054	Human genome-deriv
31	20	0.7	548	20	AAZ08316	Human lung tumour
32	20	0.7	548	21	AAK79168	Human lung tumour
33	20	0.7	548	23	AAK23243	Human lung tumour
34	20	0.7	577	24	ABL83527	Human ovarian can
35	20	0.7	638	21	AAK38427	Arabidopsis thalia
36	20	0.7	954	22	AAK92514	Human T2R14 nucleo
37	20	0.7	1287	23	AAK51375	Enterococcus faeca
38	20	0.7	1299	23	AAK53120	Enterococcus faeca
39	20	0.7	1635	20	AAK13558	Enterococcus faeca
40	20	0.7	1635	24	ABS99353	Enterococcus faeca
41	20	0.7	1817	23	AAK99001	DNA encoding novel
42	20	0.7	2493	24	ABK34875	Human CDNA encodin
43	20	0.7	2706	23	ABL16061	Drosophila melanog
44	20	0.7	4757	23	ABL16060	Drosophila melanog
45	20	0.7	4453	20	AAK23519	Human kidney amino
46	20	0.7	267156	24	ABK68560	Kidney cancer rela
47	19	0.7	330	22	AAK164817	Human polynucleoti
48	19	0.7	380	22	AAK33364	Human immune/haema
49	19	0.7	380	22	AAK93365	Human immune/haema
50	19	0.7	387	22	AAK62758	Human secreted pro
51	19	0.7	484	21	AAK34237	Drosophila melanog
52	19	0.7	573	23	ABL18649	Human nervous syst
53	19	0.7	834	22	AAK20496	Human CDNA clone
54	19	0.7	902	22	AAK06575	P3 gene of Haemoph
55	19	0.7	1023	11	AAK06119	P3 gene of Haemoph
56	19	0.7	1072	16	AAK84357	Ndel-BamHI fragmen
57	19	0.7	1074	16	AAK84356	Ndel-BamHI fragmen
58	19	0.7	1137	16	AAK84355	BamHI-XhoI fragmen
59	19	0.7	1140	11	AAK05372	P3 gene of Haemoph
60	19	0.7	1143	24	ABK66364	Streptococcus poly
61	19	0.7	1146	24	AAK46717	Group B Streptococ
62	19	0.7	1174	22	AAK23110	DNA encoding novel
63	19	0.7	1302	22	AAK46059	Human DNA encoding
64	19	0.7	1300	25	AAK57827	Human PRO3572 cDNA
65	19	0.7	1300	25	ABK59827	Human CDNA encodin
66	19	0.7	1300	25	ABK59827	Novel human secret
67	19	0.7	1300	25	AAK05834	Human secreted/tra
68	19	0.7	1300	25	ABK97878	Human PRO polynuci
69	19	0.7	1300	25	ABK78662	Human PRO polynuci
70	19	0.7	1300	25	ABK75675	Human CDNA encodin
71	19	0.7	1300	25	ABK76880	Human PRO polynuci
72	19	0.7	1300	25	ABK76880	Human CDNA encodin
73	19	0.7	1312	10	AAK90033	Haemophilus influe
74	19	0.7	1363	24	ABZ16067	Arabidopsis thalia
75	19	0.7	1477	16	AAK84354	Sall-Sall fragment
76	19	0.7	1744	22	AAK20240	Human nervous syst
77	19	0.7	2182	22	AAH16271	Human CDNA sequen
78	19	0.7	2341	24	ABK94897	Human DNA sequence
79	19	0.7	2458	23	ABL19290	Drosophila melanog
80	19	0.7	2573	23	ABL18648	Drosophila melanog
81	19	0.7	2575	11	AAK04026	Encodase



JOURNAL Genome Res. 10 (7), 939-949 (2000)  
 MEDLINE 20359837  
 PUBMED 10899143  
 REFERENCE 3 (bases 1 to 681)  
 AUTHORS Genoscope.  
 TITLE Direct Submission  
 JOURNAL Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :  
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
 - Web : www.genoscope.cns.fr)  
 COMMENT This sequence is a single read and was generated as part of a large  
 scale clone-end sequencing project of the tetraodon nigroviridis  
 genome. For more information, please take a look at  
 http://www.genoscope.cns.fr/Tetraodon.  
 FEATURES  
 source  
 1. .681  
 /organism="Tetraodon nigroviridis"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9983"  
 /clone="144M06"  
 /clone\_lib="G"  
 /note="Genoscope sequence ID : COAG144BG03Lp1-end : T7"  
 BASE COUNT 128 a 181 c 188 g 183 t 1 others  
 ORIGIN  
 Query Match 0.8%; Score 22; DB 29; Length 681;  
 Best Local Similarity 100.0%; Pred.No.31;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1763 CCCCTGGTGCTCCGGGCGAGGA 1784  
 |||||  
 Db 308 CCCCTGGTGCTCCGGGCGAGGA 329

Search completed: October 3, 2003, 04:32:35  
 Job time : 5327 secs

```

RESULT 43
AV592493/c
LOCUS
DEFINITION AV592493 Bos taurus cartilage fetus Bos taurus CDNA clone
E1CA001H05 3', mRNA sequence.
ACCESSION AV592493
VERSION AV592493
KEYWORDS EST.
SOURCE Bos taurus (cow)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE
1 (bases 1 to 471)
Takauega,A., Hirotsune,S., Itoh,R., Jitohzono,A., Suzuki,H., Aso,H.
and Sugimoto,Y.
Establishment of a high throughput EST sequencing system using
poly(A) tail-removed cDNA libraries and determination of 36,000
bovine ESTs
Nucleic Acids Res. 29 (22), E108 (2001)
21570554
1171328
COMMENT Contact: Yoshikazu Sugimoto
Animal Genetics Division
Shirakawa Institute of Animal Genetics
Odakura, Nishigo, Nishi-shirakawa, Fukushima 961-8061, Japan
Tel: 81-248-25-5641
Fax: 81-248-25-5725
Email: kazusugi@ccoc.ocn.ne.jp
Single pass sequencing.
This clone was obtained from a polyA-deleted cDNA library.
FEATURES
source
1..471
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/clone="E1CA001H05"
/tissue_type="cartilage"
/dev_stage="fetus"
/lab_host="DH10B"
/clone_lib="Bos taurus cartilage fetus"
/notes="vector: pZLI; Site_1: SalI; Site_2: NotI; Poly A
was deleted from a NotI site"
BASE COUNT 78 a 155 c 121 g 113 t
ORIGIN
Query Match 0.8%; Score 22; DB 9; Length 471;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Cy 2659 GGTCTCTCTGTCCTCCCGAGTC 2680
|||||
Db 169 GGTCTCTCTGTCCTCCCGAGTC 148

RESULT 44
BH882288
LOCUS
DEFINITION BH882288.1 WGS-Zmaysf (JM107 adapted methyl filtered) Zea mays
genomic clone hw36d12 5', genomic survey sequence.
ACCESSION BH882288
VERSION BH882288.1
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
Clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 485)
Rabinowicz,P.D., O'Shaughnessy,A.L., Balija,V., Dedhia,N.,
Katzenburger,F., King,L., Miller,B., Muller,S., Nascimento,L.,

```

```

Zutavern,T., McCombie,W.R. and Martienssen,R.A.
Genomic shotgun sequences from Zea mays (methyl-filtered)
Unpublished
Contact: W. Richard McCombie
Lita Annenberg Hazen Genome Sequencing Center
Cold Spring Harbor Laboratory
PO Box 100, Cold Spring Harbor, NY 11724, USA
Tel: 516 367 8884
Fax: 516 367 8874
Email: mcombie@cshl.org
Plate: hw36 row: d column: 12
Seq primer: -21M13UnivFwd
Class: shotgun
High quality sequence stop: 485.
FEATURES
Location/Qualifiers
1..485
/organism="Zea mays"
/mol_type="genomic DNA"
/cultivar="B73"
/db_xref="taxon:4577"
/clone="hw36d12"
/lab_host="JM107 or DH5a"
/clone_lib="WGS-Zmaysf (JM107 adapted methyl filtered)"
/notes="Organ: immature ears; Site_1: Xba I; Site_2: Xba I;
The vector was digested with XbaI and one nucleotide was
added by fill in in the recessive 3' end. The genomic DNA
was reutilized, end repaired, adaptor ligated and size
fractionated using sephadex. The resulting fragments were
between 0.8 and 3 kb and were cloned into the vector
::x/y reads in M13mp19, b/g reads in pUC19. The same
ligation was transformed in either JM107 or DH5a."
BASE COUNT 188 a 77 c 86 g 134 t
ORIGIN
Query Match 0.8%; Score 22; DB 28; Length 485;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Cy 753 AAAAAGATATAAAGGATATGAA 774
|||||
Db 226 AAAAAGATATAAAGGATATGAA 247

RESULT 45
CNS02JXM
LOCUS
DEFINITION CNS02JXM 681 bp DNA linear GSS 01-SEP-2000
Tetraodon nigroviridis genome survey sequence 17 end of clone
194M06 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION AL200803
VERSION AL200803.1
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontoidea; Tetraodontidae; Tetraodon.
1
Roest Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
Estimate of human gene number provided by genome-wide analysis
using Tetraodon nigroviridis DNA sequence
Nat. Genet. 25 (2), 235-238 (2000)
20296633
10835645
REFERENCE
2
Roest Crolius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
Saurin,W., Bernot,A. and Weissenbach,J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis

```

MEDLINE  
21570554  
11713328  
PUBMED

COMMENT  
Contact: Yoshikazu Sugimoto  
Animal Genetics Division

Shirakawa Institute of Animal Genetics  
Odakura, Nishigo, Nishi-shirakawa, Fukushima 961-8061, Japan  
Tel: 81-248-25-5641  
Fax: 81-248-25-5725

Email: kazusugi@coea.ocn.ne.jp

Single pass sequencing

This clone was obtained from a polyA-deleted cDNA library.

## FEATURES

source

1. .440  
/organism="Bos taurus"  
/mol\_type="mRNA"  
/db\_xref="taxon:9913"  
/clone="E1AD003D11"  
/cell\_type="an adipocyte cell line"  
/lab\_host="DH10B"  
/clone\_lib="Bos taurus adipocyte cell line"  
/note="Vector: pZ11; Site 1: SalI; Site 2: NotI; Poly A  
was deleted from a NotI site"

BASE COUNT 80 a 148 c 102 g 110 t

## ORIGIN

Query Match 0.8%; Score 22; DB 9; Length 440;  
Best Local Similarity 100.0%; Pred.No. 27;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2659 GGTCTCTCTGTCCTCCCGAGTC 2680

Db 383 GGTCTCTCTGTCCTCCCGAGTC 362

## RESULT 41

BH882289/c

LOCUS

DEFINITION hw36d12.g1 WGS-Zmayf (JM107 adapted methyl filtered) Zea mays  
genomic clone hw36d12 5', genomic survey sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.

REFERENCE

AUTHORS

Rabinowicz, P.D., O'Shaughnessy, A.L., Balija, V., Dedhia, N.,  
Katzenburger, F., King, L., Miller, B., Muller, S., Nascimento, L.,  
Zutavern, T., McCombie, W.R., and Martienssen, R.A.

Genomic shotgun sequences from Zea mays (methyl-filtered);

Unpublished

JOURNAL

COMMENT

Contact: W. Richard McCombie  
Lita Annenberg Hazen Genome Sequencing Center  
Cold Spring Harbor Laboratory  
PO Box 100, Cold Spring Harbor, NY 11724, USA  
Tel: 516 367 8884  
Fax: 516 367 8874

Email: mcombie@cshl.org

Plate: hw36 row: d column: 12

Seq primer: -21M13UnivFwd

Class: shotgun

High quality sequence stop: 487.

## FEATURES

source

1. .450  
/organism="Zea mays"  
/mol\_type="genomic DNA"  
/cultiivar="B73"  
/db\_xref="taxon:4577"  
/clone="hw36d12"  
/lab\_host="JM107 or DH5a"  
/clone\_lib="WGS-Zmayf (JM107 adapted methyl filtered)"

/note="Organ: immature ears; Site 1: Xba I; Site 2: Xba I;  
The vector was digested with XbaI and one nucleotide was  
added by fill in in the recessive 3' end. The genomic DNA  
was rebligated, end repaired, adaptor ligated and size  
fractionated using sephadex. The resulting fragments were  
between 0.8 and 3 kb and were cloned into the vector  
(-x/y reads in M13p19, -b/g reads in pUC19). The same  
ligation was transformed in either JM107 or DH5a."

BASE COUNT 119 a 79 c 70 g 182 t

## ORIGIN

Query Match 0.8%; Score 22; DB 28; Length 450;  
Best Local Similarity 100.0%; Pred.No. 28;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 753 AAAAAGATATAAGGATATGAA 774

Db 245 AAAAAGATATAAGGATATGAA 224

## RESULT 42

AV599587/c

LOCUS

DEFINITION 454 bp mRNA linear EST 27-NOV-2001  
AV599587 Bos taurus cartilage fetus Bos taurus cDNA clone  
E1CA044H10 3', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Bos taurus (cow)  
Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovidae; Bovinae; Bos.

REFERENCE

AUTHORS

Takasuga, A., Hirotsune, S., Itoh, R., Jitchazon, A., Suzuki, H., Aso, H.  
and Sugimoto, Y.  
Establishment of a high throughput EST sequencing system using  
poly(A) tail-removed cDNA libraries and determination of 36,000  
bovine ESTs

JOURNAL

MEDLINE

PUBMED

COMMENT

Contact: Yoshikazu Sugimoto

Animal Genetics Division

Shirakawa Institute of Animal Genetics

Odakura, Nishigo, Nishi-shirakawa, Fukushima 961-8061, Japan

Tel: 81-248-25-5641

Fax: 81-248-25-5725

Email: kazusugi@coea.ocn.ne.jp

Single pass sequencing.

This clone was obtained from a polyA-deleted cDNA library.

## FEATURES

source

1. .454  
/organism="Bos taurus"  
/mol\_type="mRNA"  
/db\_xref="taxon:9913"  
/clone="E1CA044H10"  
/tissue\_type="cartilage"  
/dev\_stage="fetus"  
/lab\_host="DH10B"  
/clone\_lib="Bos taurus cartilage fetus"  
/note="Vector: pZ11; Site 1: SalI; Site 2: NotI; Poly A  
was deleted from a NotI site"

BASE COUNT 78 a 154 c 115 g 107 t

## ORIGIN

Query Match 0.8%; Score 22; DB 9; Length 454;  
Best Local Similarity 100.0%; Pred.No. 28;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2659 GGTCTCTCTGTCCTCCCGAGTC 2680

Db 182 GGTCTCTCTGTCCTCCCGAGTC 161

Email: kazusugi@cocoa.ocn.ne.jp  
Single pass sequencing.  
This clone was obtained from a polyA-deleted cDNA library.

FEATURES  
source Location/Qualifiers

1..367  
/organism="Bos taurus"  
/mol\_type="mRNA"  
/db\_xref="taxon:9913"  
/clone="E1CA023A01"  
/tissue\_type="cartilage"  
/dev\_stage="fetus"  
/lab\_host="DH10B"  
/clone\_lib="Bos taurus cartilage fetus"  
/note="Vector: pZLI; Site\_1: SalI; Site\_2: NotI; Poly A  
was deleted from a NotI site"

BASE COUNT 65 a 124 c 88 g 89 t : others  
ORIGIN

Query Match 0.8%; Score 22; DB 9; Length 367;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2659 GGTCTCTCTGGTCCCGGAGTC 2680

Db 191 GGTCTCTCTGGTCCCGGAGTC 160

RESULT 38

CB801379

LOCUS CB801379 412 bp mRNA linear EST 16-MAY-2003  
DEFINITION AMGNNUC:MRPE4-00162-D9-A mrpe4 (10380) Rattus norvegicus cDNA clone  
mrpe4-00162-d9 5', mRNA sequence.

ACCESSION CB801379

VERSION CB801379.1 GI:29914582

KEYWORDS EST.

SOURCE Rattus norvegicus (Norway rat)

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

REFERENCE 1 (bases 1 to 412)

Amgen EST Program.

TITLE Amgen Rat EST Program

JOURNAL Unpublished

COMMENT Contact: Dan Fitzpatrick

Amgen, Inc

One Amgen Center Drive, Thousand Oaks, CA 91320-1799, USA

Tel: 805 447-4881

Plate: 00162 row: d column: 9.

FEATURES  
source Location/Qualifiers

1..412  
/organism="Rattus norvegicus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10116"  
/clone="mrpe4-00162-d9"  
/tissue\_type="placenta embryo"  
/clone\_lib="mrpe4 (10380)"  
/note="Vector: pSPOR1; Site\_1: SalI; Site\_2: NotI;  
placenta embryo day 17"

BASE COUNT 83 a 122 c 147 g 60 t

ORIGIN

Query Match

Best Local Similarity 0.8%; Score 22; DB 14; Length 412;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2702 GCAAGAAGGCTCTCCAGGAGA 2723

Db 315 GCAAGAAGGCTCTCCAGGAGA 336

RESULT 39

AV597967/c

LOCUS AV597967 424 bp mRNA linear EST 27-NOV-2001  
DEFINITION E1CA035B11 3', mRNA sequence.  
Bos taurus cDNA clone

ACCESSION AV597967

VERSION AV597967.1 GI:9715461

KEYWORDS EST.

SOURCE Bos taurus (cow)

ORGANISM Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovidae; Bovinae; Bos.

REFERENCE 1 (bases 1 to 424)

Takasuga,A., Hirotsune,S., Itoh,R., Jitohzono,A., Suzuki,H., Aso,H.

and Sugimoto,Y.

Establishment of a high throughput EST sequencing system using

poly(A) tail-removed cDNA libraries and determination of 36,000

bovine ESTs

Nucleic Acids Res. 29 (22), E108 (2001);

21570554

PMID 11713328

COMMENT

Contact: Yoshikazu Sugimoto

Animal Genetics Division

Shirakawa Institute of Animal Genetics

Odakura, Nishigo, Nishi-Shirakawa, Fukushima 961-8061, Japan

Tel: 81-248-25-5641

Fax: 81-248-25-5725

Email: kazusugi@cocoa.ocn.ne.jp

Single pass sequencing.

This clone was obtained from a polyA-deleted cDNA library.

FEATURES  
source Location/Qualifiers

1..424

/organism="Bos taurus"

/mol\_type="mRNA"

/db\_xref="taxon:9913"

/clone="E1CA035B11"

/tissue\_type="cartilage"

/dev\_stage="fetus"

/lab\_host="DH10B"

/clone\_lib="Bos taurus cartilage fetus"

/note="Vector: pZLI; Site\_1: SalI; Site\_2: NotI; Poly A

was deleted from a NotI site"

BASE COUNT 74 a 144 c 109 g 96 t 1 others

ORIGIN

Query Match 0.8%; Score 22; DB 9; Length 424;

Best Local Similarity 100.0%; Pred. No. 27;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2659 GGTCTCTCTGGTCCCGGAGTC 2680

Db 185 GGTCTCTCTGGTCCCGGAGTC 164

RESULT 40

AV614172/c

LOCUS AV614172 440 bp mRNA linear EST 28-NOV-2001

DEFINITION AV614172 Bos taurus adipocyte cell line Bos taurus cDNA clone

E1AD003D11 3', mRNA sequence.

ACCESSION AV614172

VERSION AV614172.1 GI:9749842

KEYWORDS EST.

SOURCE Bos taurus (cow)

ORGANISM Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

Bovidae; Bovinae; Bos.

REFERENCE 1 (bases 1 to 440)

Takasuga,A., Hirotsune,S., Itoh,R., Jitohzono,A., Suzuki,H., Aso,H.

and Sugimoto,Y.

Establishment of a high throughput EST sequencing system using

poly(A) tail-removed cDNA libraries and determination of 36,000

bovine ESTs

Nucleic Acids Res. 29 (22), E108 (2001);

JOURNAL

QY 1405 CCTGGTCAAGATGTAAGCTCGGATATCAGGAAT 1439  
 |||||  
 Db 346 CCTGGTCAAGATGTAAGCTCGGATATCAGGAAT 380  
 |||||

RESULT 35  
 AQ514534 556 bp DNA linear GSS 05-MAY-1999  
 LOCUS HS 5081.B1.A01.SP6E.RPCL1-11 Human Male BAC Library Homo sapiens  
 DEFINITION genomic\_clone Plate=657 Col=1 Row=B, genomic survey sequence.  
 ACCESSION AQ514534  
 VERSION AQ514534.1 GI:4742898  
 KEYWORDS GSS.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 556)  
 AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
 Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and  
 Hood,L.  
 TITLE Sequence-tagged connectors: A sequence approach to mapping and  
 scanning the human genome  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)  
 MEDLINE 9380589  
 PUBMED 10449764  
 COMMENT Contact: Mahairas GG, Wallace JC, Hood L  
 High Throughput Sequencing Center  
 University of Washington  
 401 Queen Anne Avenue North, Seattle, WA 98109, USA  
 Tel: (206) 616-3618  
 Fax: (206) 616-3887  
 Email: jwallace@u.washington.edu  
 Clones are derived from the human BAC library RPCL1-11. For BAC  
 library availability, please contact Pieter de Jong  
 (pieterdejong.med@u.washington.edu). Clones may be purchased from  
 BACPAC Resources ([http://bacpac.med.buffalo.edu/ordering\\_bac.htm](http://bacpac.med.buffalo.edu/ordering_bac.htm))  
 or from Resear h Genetics (<http://www.htsc.washington.edu>)  
 Plate: 657 row: B column: 1  
 Seq primer: SP6  
 Class: BAC ends  
 High quality sequence stop: 556.  
 Location/Qualifiers  
 1..556  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /clone="Plate=657 Col=1 Row=B"  
 /sex="male"  
 /clone\_lib="RPCL1-11 Human Male BAC Library"  
 /note="vector: pBACe3.6; Site 1: EcoRI; Site 2: EcoRI;  
 Male blood DNA was isolated from one randomly chosen donor  
 and partially digested with a combination of EcoRI and  
 EcoRI Methylase. Size selected DNA was cloned into the  
 pBACe3.6 vector at EcoRI sites"  
 BASE COUNT 191 a 106 c 103 g 151 t 5 others  
 ORIGIN

Query Match 0.8%; Score 23; DB 28; Length 556;  
 Best Local Similarity 100.0%; Pred. No. 9;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1833 ATTAATGGCCAAAGGAGAAA 1855  
 |||||  
 Db 367 ATTAATGGCCAAAGGAGAAA 389  
 |||||

RESULT 36  
 CC086742 980 bp DNA linear GSS 16-APR-2003  
 LOCUS CSU-K33r.15N22.SP6 CSU-K33r genomic clone  
 DEFINITION

CSU-K33r.15N22, genomic survey sequence.  
 CC086742  
 VERSION CS086742.1 GI:29940194  
 KEYWORDS GSS.  
 SOURCE Aedes aegypti (yellow fever mosquito)  
 ORGANISM Aedes aegypti  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Aedes.  
 REFERENCE 1 (bases 1 to 980)  
 AUTHORS Loftus,B., Shetty,J., Seversen,D., Brown,S. and Knudson,D.  
 JOURNAL End sequencing of Aedes aegypti BACs  
 COMMENT Unpublished  
 Other\_GSS: CSU-K33r.15N22.T7  
 Contact: Brendan Loftus  
 Department of Eukaryotic Genomics  
 TIGR  
 9712 Medical Center Drive, Rockville, MD 20850, USA  
 Tel: 301-838-3543  
 Fax: 301-838-0208  
 Email: entas@tigr.org  
 Library was provided by Susan Brown and Dennis Knudson at Colorad  
 State University.  
 Seq primer: SP6  
 Class: BAC ends.  
 Location/Qualifiers  
 1..980  
 /organism="Aedes aegypti"  
 /mol\_type="genomic DNA"  
 /strain="Rexville"  
 /db\_xref="taxon:7159"  
 /clone="CSU-K33r.15N22"  
 /clone\_lib="CSU-K33r"  
 /note="vector: pBeloBAC11; Site 1: HindIII"  
 BASE COUNT 305 a 198 c 204 g 273 t  
 ORIGIN

Query Match 0.8%; Score 23; DB 29; Length 980;  
 Best Local Similarity 100.0%; Pred. No. 10;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 429 GGATGGCAATCCCAAGATGACG 451  
 |||||  
 Db 567 GGATGGCAATCCCAAGATGACG 945  
 |||||

RESULT 37  
 AV595980/c  
 LOCUS AV595980 Bos taurus cartilage fetus Bos taurus cDNA clone  
 DEFINITION E1CA02JA01.3', mRNA sequence.  
 ACCESSION AV595980  
 VERSION AV595980.1 GI:9712989  
 KEYWORDS EST.  
 SOURCE Bos taurus (cow)  
 ORGANISM Bos taurus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 Bovidae; Bovinae; Bos.  
 REFERENCE 1 (bases 1 to 367)  
 AUTHORS Takasuga,A., Hitotsune,S., Itoh,R., Jitohzono,A., Suzuki,H., Aso,H.  
 and Sugimoto,Y.  
 TITLE Establishment of a high throughput EST sequencing system using  
 poly(A) tail-removed cDNA libraries and determination of 36,000  
 bovine ESTs  
 JOURNAL Nucleic Acids Res. 29 (22), E108 (2001)  
 MEDLINE 21570554  
 PUBMED 11713328  
 COMMENT Contact: Yoshikazu Sugimoto  
 Animal Genetics Division  
 Shirakawa Institute of Animal Genetics  
 Odakura, Nishigo, Nishi-shirakawa, Fukushima 961-8061, Japan  
 Tel: 81-248-25-5641  
 Fax: 81-248-25-5725

```

/db_xref="taxon:9606"
/clone="Plate=1003 Col=13 Row=F"
/sex="male"
/clone_lib="RPC1-11 Human Male BAC Library"
/note="vector: pBACE3.6; Site 1: EcoRI; Site 2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBACE3.6 vector at EcoRI sites"
BASE COUNT      217 a      124 c      96 g      213 t      18 others
ORIGIN
Query Match      1.3%; Score 36; DB 28; Length 668;
Best Local Similarity 100.0%; Pred.No. 2.2e-06;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1135 TTAGGATCTTGATCAATGGCGCAACCAACCAATTGGA :173
|||
DB 323 TTAGGATCTTGATCAATGGCGCAACCAACCAATTGGA 288

RESULT 33
BM724875      691 bp      mRNA      linear      EST 01-MAR-2002
LOCUS
DEFINITION
UI-E-E01-aja-a-05-0-UI-r1 UI-E-E01 Homo sapiens cDNA clone
ACCESSION
BM724875
VERSION
BM724875.1 GI:19046206
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 691)
AUTHORS
Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE
Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL
Genome Res. 6 (9), 791-806 (1996)
MEDLINE
97044477
PUBMED
8889548
COMMENT
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
Seq primer: M13 Reverse.
FEATURES
Location/Qualifiers
1..691
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-E-E01-aja-a-05-0-UI"
/tissue_type="fetal eye"
/dev_stage="fetal"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-E-E01"
/note="Organ: eye; Vector: p773-Pac (Pharmacia) with a
modified polylinker; Site 1: EcoRI; Site 2: Not I;
UI-E-E01 is a normalized cDNA library containing the
following tissue(s): fetal eye. The library was
constructed according to Bonaldo, Lennon and Soares.
Genome Research, 6:791-806, 1996. First strand cDNA
synthesis was primed with an oligo-dT primer containing a
Not I site. Double stranded cDNA was ligated to an EcoRI
adaptor, digested with Not I, and cloned directionally

```

```

into p773-Pac vector. The oligonucleotide used to prime
the synthesis of first-strand cDNA contains a library tag
sequence that is located between the Not I site and the
(dT)18 tail. The sequence tag for this library is
CGCGTATACC. This library was created for the program, Gene
Discovery in the Visual System, supported by National Eye
Institute (NEI)."
BASE COUNT      206 a      120 c      131 g      233 t      1 others
ORIGIN
Query Match      1.3%; Score 36; DB 12; Length 651;
Best Local Similarity 100.0%; Pred.No. 2.3e-06;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2830 AGAAGAGATCCGTTTCAGAAAGGACCAACTATTAG 2865
|||
DB 1 AGAAGAGATCCGTTTCAGAAAGGACCAACTATTAG 36

RESULT 34
BE899929      331 bp      mRNA      linear      EST 25-APR-2001
LOCUS
DEFINITION
180456 MARC 4BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION
BE899929
VERSION
BE899929.1 GI:10387679
KEYWORDS
EST.
SOURCE
Bos taurus (cow)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
Bovidae; Bovinae; Bos.
1 (bases 1 to 331)
AUTHORS
Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.C., Stone,R.T.,
Casas,E., Wray,J.E., White,J., Cho,S., Fahrnkrug,S.C., Bennett
,G.L., Heaton,M.P., Laegreid,W.M., Rohrer,G.A., Chitko-McKown,C.G.,
Perteza,G., Holt,I., Karamycheva,S., Jiang,F., Quackenbush,J. and
Keefe,J.W.
TITLE
Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
JOURNAL
Genome Res. 11 (4), 626-630 (2001)
MEDLINE
21180013
PUBMED
11282978
COMMENT
Contact: Smith TP
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt. trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCTCAGTCACG
Plate: 85 row: E column: 7
Seq primer: ATTTAGGTGACACTATAG.
Location/Qualifiers
1..531
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="MARC 4BOV"
/note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;
Library made from pooled tissue from day 20 and day 40
embryos."
BASE COUNT      161 a      120 c      146 g      104 t
ORIGIN
Query Match      1.2%; Score 35; DB 10; Length 531;
Best Local Similarity 100.0%; Pred.No. 6.9e-06;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

AQ356449/c  
 LOCUS BX391778 778 bp DNA linear GSS 24-JAN-1999  
 DEFINITION CITBI-E1-2541D7.TF CITBI-E1 Homo sapiens genomic clone 2541D7,  
 genomic survey sequence.  
 ACCESSION AQ356449  
 VERSION AQ356449.1 GI:4183622  
 KEYWORDS GSS.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 778)  
 AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., Shizuya,H., Simon,M. and  
 Venter,J.C.  
 TITLE Use of BAC End Sequences from CalTech Libraries for Sequence-Ready  
 Map Building  
 JOURNAL Unpublished  
 COMMENT Contact: Shaying Zhao, William Nierman, Mark Adams  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: hbe@tigr.org  
 Clones are available from Research Genetics (info@resgen.com). BAC  
 end search page:  
 http://www.tigr.org/tdb/humgen/bac\_end\_search/bac\_end\_search.htm.  
 Seq primer: M13-21  
 Class: BAC ends.

## FEATURES

source

1..778  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /clone="2541D7"  
 /sex="male"  
 /cell\_type="sperm"  
 /clone\_lib="CITBI-E1"  
 /note="Vector: pBelBAC11; Site\_1: EcoRI; Site\_2: EcoRI;  
 Caltech Human BAC Library D"  
 BASE COUNT 289 a 109 c 118 g 262 t  
 ORIGIN

Query Match 1.5%; Score 42; DB 28; Length 778;  
 Best Local Similarity 100.0%; Pred.No. 2e-09;  
 Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGGCTACTATATACATTCTCTGCATGTTTGGTGCTG 42  
 |||||  
 Db 61 ATGGCTACTATATACATTCTCTGCATGTTTGGTGCTG 20  
 |||||

RESULT 31  
 BX391778/c  
 LOCUS BX391778 883 bp mRNA linear EST 13-MAY-2003  
 DEFINITION Clone CSODI071YH18 3-PRIME, mRNA sequence.  
 ACCESSION BX391778  
 VERSION BX391778.1 GI:30607719  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 883)  
 AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished  
 COMMENT Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 EVRY cedex - France  
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
 Library was constructed by Life Technologies, a division of

Invitrogen. This sequence belongs to sequence cluster 3377.r For  
 more information about this cluster, see  
 http://www.genoscope.cns.fr/  
 cgi-bin/cluster.cgi?seq=CS0BAI027ZH02\_CS02512\_1&cluster=3377.r.  
 Contact : Feng Liang Email : fliang@lifetech.com URL :  
 http://fulllength.invitrogen.com/InvitrogenCorporation.1600  
 Faraday Avenue Genoscope sequence ID : CS0BAI027ZH02\_CS02512\_1.

## FEATURES

source

1..883  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0DI071YH18"  
 /tissue\_type="PLACENTA COT 25-NORMALIZED"  
 /clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"  
 /note="1st strand cDNA was primed with a NotI-oligo (dT)  
 primer. Five prime end enriched double-strand cDNA was  
 digested with NotI and cloned into the NotI and EcoR V  
 sites of the pCMVSPORT 6 vector. Library was normalized."  
 BASE COUNT 274 a 166 c 163 g 275 t  
 ORIGIN

Query Match 1.3%; Score 38; DB 13; Length 883;  
 Best Local Similarity 100.0%; Pred.No. 2.3e-07;  
 Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2828 CCAGAAGAGATCCGTTTCAGAAAAGGACCAACTATTAG 2865  
 o |||||  
 Db 777 CCAGAAGAGATCCGTTTCAGAAAAGGACCAACTATTAG 740  
 |||||

RESULT 32  
 AQ836646/c  
 LOCUS AQ836646 668 bp DNA linear GSS 30-AUG-1999  
 DEFINITION HS 5427 B1-C07 77A RPCI-11 Human Male BAC Library Homo sapiens  
 genomic clone Plate=i003 Col=13 Row=F, genomic survey sequence.  
 ACCESSION AQ836646  
 VERSION AQ836646.1 GI:5906520  
 KEYWORDS GSS.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 668)  
 AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
 Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and  
 Hood,L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and  
 scanning the human genome  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)  
 MEDLINE 99380589  
 PUBMED 10449764  
 COMMENT Contact: Mahairas GG, Wallace JC, Hood L  
 High Throughput Sequencing Center  
 University of Washington  
 421 Queen Anne Avenue North, Seattle, WA 98105, USA  
 Tel: (206) 616-3618  
 Fax: (206) 616-3887  
 Email: jwallace@u.washington.edu  
 Clones are derived from the human BAC library RPCI-11. For BAC  
 library availability, please contact Pieter de Jong  
 (pieter@dejong.med.bufo.edu). Clones may be purchased from  
 BACPAC Resources (http://bacpac.med.bufo.edu/ordering\_bac.htm)  
 or from Resear h Genetics (info@resgen.com). BAC end Web Server:  
 http://www.htsc.washington.edu  
 Plate: 1003 row: F column: 13  
 Seq primer: 7  
 Class: BAC ends

High quality sequence stop: 668.  
 Location/Qualifiers  
 1..668  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKF2p686B2113"
/tissue_type="human skeletal muscle"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="686 (synonym: blcc3)"
/notes="vector: pTriplex2; Site_1: SfiI; Site_2: SfiI;
cDNA-collection"
BASE COUNT 193 a 77 c 107 g 223 t
ORIGIN
Query Match 1.8%; Score 51; DB 9; Length 600;
Best Local Similarity 100.0%; Pred. No. 4.9e-14;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1804 GGAATCCAGGATTTCTGGAACCGAGGATTAATGGCCAAAGGGAGAA 1854
|||||
Db 532 GGAATCCAGGATTTCTGGAACCGAGGATTAATGGCCAAAGGGAGAA 582
|||||

RESULT 28
BE236040 518 bp mRNA linear EST 10-JUL-2000
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Sus scrofa (pig)
Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE
1 (bases 1 to 518)
AUTHORS
Fahrenkrug,S.C., Smith,T.P.L., Preking,B.A., Cho,J., White,J.,
Vallet,J., Wise,T., Rohrer,G.A., Perle,G., Sultana,R., Quackenbush
,J. and Keele,J.W.
Pig gene discovery by normalized cDNA-library sequencing and
EST cluster assembly
Mamm. Genome 13 (8), 475-478 (2002)
22213789
2226715
CONTACT: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@mail.marc.usda.gov
Single pass sequencing. Bases called and alt trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -mismatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCTCCGTCAGCAGC
Plate: 85 row: G column: 4
Seq primer: ATTAGGTCACACTATAG.
FEATURES
Location/Qualifiers
1..518
/organism="Sus scrofa"
/mol_type="mRNA"
/db_xref="taxon:9823"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="MARC 1P1G"
/notes="vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
Library made from pooled tissue from day 11, 13, 15, 20,
and 30 embryos."
BASE COUNT 157 a 105 c 175 g 81 t
ORIGIN
Query Match 1.6%; Score 45; DB 10; Length 518;
Best Local Similarity 100.0%; Pred. No. 5.4e-11;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1692 TCTCGAGAACCCAGGACACATCGGAAGGATCGGATTAATGGGTAG 1736
|||||
Db 111 TCTCGAGAACCCAGGACACATCGGAAGGATCGGATTAATGGGTAG 155
|||||

RESULT 29
R77744 372 bp mRNA linear EST 07-JUN-1995
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 372)
AUTHORS
Hilfier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
,M., Huitman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
,R., Williamson,A., Wohlmann,P. and Wilson,R.
The WashU-Merck EST Project
Unpublished
CONTACT: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 850; St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1820
Email: est@watson.wustl.edu
Insert Size: 1108
High quality sequence stops: 328
Source: IVAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Possible reversed clone: polyt not found
Insert Length: 1108 Std Error: 0.00
Seq primer: Promega -21m13
High quality sequence stop: 328.
FEATURES
Location/Qualifiers
1..372
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:566422"
/db_xref="taxon:9606"
/clone="IMAGE:145586"
/sex="Female"
/dev_stage="placenta obtained at birth (full term)"
/clone_lib="Soares placenta NBZHP"
/notes="Organ: placenta; Vector: p7T73D (Pharmacia) with a
modified polylinker; Site_1: NotI; Site_2: Eco RI; 1st
strand cDNA was primed with a NotI - oligo(dT) primer (5'
AACTGGAGCAATTCGGCCGAGCAATTTTCTTTTCTTTT 3').
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with NotI and cloned into the NotI
and Eco RI sites of the modified p7T73 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo. "
BASE COUNT 113 a 88 c 52 g 115 t 4 others
ORIGIN
Query Match 1.5%; Score 44; DB 14; Length 372;
Best Local Similarity 100.0%; Pred. No. 1.6e-10;
Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1576 CCAGGATCAAGGTCGAATGGTCCCAAGGAGACCAAGGATC 1619
|||||
Db 173 CCAGGATCAAGGTCGAATGGTCCCAAGGAGACCAAGGATC 130
|||||

RESULT 30

```



adaptor-primer. EcoRI adaptors were ligated, followed by digestion with XhoI, for directional cloning into predigested lambda ZAP Express."

BASE COUNT 53 a 59 c 77 g 52 t  
 ORIGIN  
 Query Match 2.0%; Score 58; DB 9; Length 241;  
 Best Local Similarity 100.0%; Pred. No. 1e-17;  
 Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2443 GATCATTGCTTCCCAACATGCTCCCGGGTATCTCTGGCCACCTGTGCCGATAG 2500  
 |||||  
 Db 1 GATCATTGCTTCCCAACATGCTCCCGGGTATCTCTGGCCACCTGTGCCGATAG 58

RESULT 25  
 CB268816 428 bp mRNA linear EST 20-FEB-2003  
 LOCUS 1007722 Human Fat Cell 5'-Stretch Plus cDNA Library Homo sapiens  
 DEFINITION CDNA 5', mRNA sequence.

ACCESSION CB268816  
 VERSION 1  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 428)  
 AUTHORS Yang, R.-Z., Shuldiner, A. and Gong, D.-W.  
 TITLE EST analysis of human adipose gene expression  
 JOURNAL Unpublished

COMMENT Contact: Gong Da-Wei  
 Division of Endocrinology, Diabetes and Nutrition  
 University of Maryland  
 660 Redwood St, HH497, Baltimore, MD 21201, USA  
 Tel: 410 706 1672  
 Fax: 410 706 1622

Email: dgong@medicine.umaryland.edu  
 PCR PRIMERS  
 FORWARD: CTCGGGAAGCGCCGATGTGTGTT  
 BACKWARD: AATACGACTACTATAGGCGAATGG  
 Seq primer: GTTGTACCGCGAATC.

FEATURES  
 source  
 1..428  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /sex="Male and Female"  
 /tissue\_type="Adipose"  
 /clone\_lib="Human Fat Cell 5'-Stretch Plus cDNA Library"  
 /note="Vector: lambdaTriplex"

BASE COUNT 122 a 83 c 68 g 135 t 20 others

ORIGIN  
 Query Match 2.0%; Score 57; DB 14; Length 428;  
 Best Local Similarity 100.0%; Pred. No. 3.9e-17;  
 Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2809 CTATGTTTGTAGTGAATTCCTCCAGAGAGATCCGTTCCAGAAAGGACCAACTATTAG 2865  
 |||||  
 Db 12 CTATGTTTGTAGTGAATTCCTCCAGAGAGATCCGTTCCAGAAAGGACCAACTATTAG 68

RESULT 26  
 BX387812 903 bp mRNA linear EST 08-MAY-2003  
 LOCUS BX387812  
 DEFINITION BX387812 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA  
 Clone CS0D1071YH18 5-PRIME, mRNA sequence.

ACCESSION BX387812  
 VERSION 1  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)

ORGANISM

REFERENCE  
 AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished

COMMENT Contact: Genoscope

Genoscope - Centre National de Sequencage  
 BP 191 91006 EVRY cedex - France  
 Email: segre@genoscope.cns.fr, Web : www.genoscope.cns.fr  
 Library was constructed by Life Technologies, a division of  
 Invitrogen. This sequence belongs to sequence cluster 3377.r For  
 more information about this cluster, see  
 http://www.genoscope.cns.fr/

cgi-bin/cluster.cgi?seq=CS0A1071DD09QPL&cluster=3377.r. Contact :  
 Peng Liang Email : fliang@lifetech.com URL :  
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
 Faraday Avenue Genoscope sequence ID : CS0A1071DD09QPL.

FEATURES  
 source

1..903  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0D1071YH18"  
 /tissue\_type="PLACENTA COT 25-NORMALIZED"  
 /clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"  
 /note="1st strand cDNA was primed with a NotI-oligo(dT)  
 primer. Five prime end enriched, double-strand cDNA was  
 digested with Not I and cloned into the Not I and EcoR V  
 sites of the pCMVSPORT 6 vector. Library was normalized."

BASE COUNT 294 a 139 c 173 g 258 t 39 others

ORIGIN  
 Query Match 2.0%; Score 56; DB 13; Length 903;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-16;  
 Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2705 AAGAAGTCTCTCCAGGAGACCCAGGCTCTCCCTGCACAGATGGAGACCTGGAAAA 2760  
 |||||  
 Db 57. AAGAAGTCTCTCCAGGAGACCCAGGCTCTCCCTGCACAGATGGAGACCTGGAAAA 626

RESULT 27  
 AL602341  
 LOCUS

DEFINITION AL602341 600 bp mRNA linear EST 14-AUG-2001

ACCESSION AL602341

VERSION AL602341.1 GI:15165847

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 600)

AUSCHERS Ansgor, W., Winkner, U., Mewes, W., Weil, B. and Wiemann, S.

TITLE EST (Ansgor, W., Winkner, U., Mewes, W., Weil, B. and Wiemann, S.)

JOURNAL Unpublished

COMMENT Contact: Ansgor W

MIPS  
 Ingolstaedter Landstr. 1, D-85764 Neuherberg, Germany  
 This is the 5' sequence of the clone insert  
 Clone from S. Wiemann, Molecular Genome Analysis, German Cancer  
 Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;  
 sequenced by EMBL (European Molecular Biology Laboratories,  
 Heidelberg/Germany) within the cDNA sequencing consortium of the  
 German Genome Project.  
 No 5' sequence available.

This clone (DKFZp666B2113) is available at the RZPD in Berlin.  
 Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059  
 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

FEATURES  
 source

1..600

Email: asimpson@ludwig.org.br  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2-CM1-NN1006-300  
 500-231-g01&t3=2000-05-30&t4=1)  
 Seq primer: puc 18 forward  
 High quality sequence start: 7  
 High quality sequence stop: 329.

## FEATURES

Location/Qualifiers  
 1..330  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /db\_xref="taxon:9606"  
 /clone\_lib="NN1006"  
 /note="Organ: Lung; Vector: p7T3-Pac (Pharmacia) with a modified pC-ylinker; Site 1: EcoR I; Site 2: Not I; UI-CF-FNO is a subtracted cDNA library derived from two normal Human lung epithelial cell libraries (EN1 and D91). The library was subtracted according to according to Bernaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. For additional information, contact: bentso-soares@ludwig.edu  
 TAG\_Lib=UI-CF-FNO  
 TAG\_Tissue=Lung Epithelial Cells Tissue nos 359-368  
 TAG\_SEQ=GGCTGTAGGC"

BASE COUNT 94 a 74 c 64 g 38 t

Query Match 2.3%; Score 66; DB 10; Length 330;  
 Best Local Similarity 100.0%; Pred. No. 9.4e-22;  
 Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2747 GAGACATGGAAACCTGGATCCAAAGGACACAGGCCCCCAGGATCTGGACCCAT 2806  
 |||||  
 DB 15 GAGACATGGAAACCTGGATCCAAAGGACACAGGCCCCCAGGATCTGGACCCAT 74

QY 2807 CACTAT 2812  
 |||||  
 DB 75 CACTAT 80

## RESULT 23

CB241974/c  
 LOCUS CB241974 246 bp mRNA linear EST 12-FEB-2003  
 DEFINITION UI-CF-FNO-aga-o-08-0-UI-s1 UI-CF-FNO Homo sapiens cDNA clone  
 UI-CF-FNO-aga-o-08-0-UI 3', mRNA sequence.

ACCESSION CB241974  
 VERSION CB241974.1 GI:28363618

KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 246)  
 AUTHORS Ronaldo, M.F., Lennon, G. and Soares, M.B.  
 TITLE Normalization and subtraction: two approaches to facilitate gene discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)  
 MEDLINE 97044477  
 PubMed 8889548

COMMENT Contact: McCray, PB  
 University of Iowa  
 2024 University of Iowa Med Labs, Iowa City, IA 52242, USA  
 Tel: 319 356 4866  
 Fax: 319 356 7171  
 Email: paul-mccray@uiowa.edu

Tissue Procurement: Dr. M. J. Welsh, University of Iowa  
 cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com) or from Open Biosystems (www.openbiosystems.com).  
 Seq primer: M13 FORWARD  
 POLYA=yes.

## FEATURES

Location/Qualifiers  
 1..246  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone\_lib="Human Lung Epithelial cells"  
 /tissue="Human Lung Epithelial cells"  
 /ab\_host="DH1CB (Life Technologies) (T1 phage resistant)"  
 /clone\_lib="UI-CF-FNO"  
 /note="Organ: Lung; Vector: p7T3-Pac (Pharmacia) with a modified pC-ylinker; Site 1: EcoR I; Site 2: Not I; UI-CF-FNO is a subtracted cDNA library derived from two normal Human lung epithelial cell libraries (EN1 and D91). The library was subtracted according to according to Bernaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. For additional information, contact: bentso-soares@ludwig.edu  
 TAG\_Lib=UI-CF-FNO  
 TAG\_Tissue=Lung Epithelial Cells Tissue nos 359-368  
 TAG\_SEQ=GGCTGTAGGC"

BASE COUNT 56 a 53 c 39 g 98 t

Query Match 2.2%; Score 64; DB 14; Length 246;  
 Best Local Similarity 100.0%; Pred. No. 9.1e-21;  
 Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1580 GATCAAGGCTGAATGGTCCAAAGGACAAAGGATCACCTGGATTTATGGCAAAA 1639  
 |||||  
 DB 219 GATCAAGGCTGAATGGTCCAAAGGACAAAGGATCACCTGGATTTATGGCAAAA 160

QY 1640 AGGG 1643  
 |||||  
 DB 159 AGGG 156

## RESULT 24

AA247907  
 LOCUS J6114.seq.F Human fetal heart, Lambda ZAP Express Homo sapiens cDNA  
 DEFINITION AA247907  
 56, mRNA sequence.

ACCESSION AA247907  
 VERSION AA247907.1 GI:1880395

KEYWORDS EST.  
 SOURCE Homo sapiens (human)

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 241)  
 AUTHORS Liew, C.C.

TITLE CDNAs from human fetal heart (1997)

JOURNAL Unpublished  
 COMMENT Contact: Liew CC  
 Brigham and Women's Hospital  
 Harvard Medical School  
 75 Francis St. Boston, MA 02115, USA

Tel: 6177328915  
 Fax: 6179750995  
 Email: clliew@rics.bwh.harvard.edu

PCR Primers  
 FORWARD: 5' OCCAAGCTCGAATTAACCTCACTAAAGG 3'  
 BACKWARD: 5' CCAGTGAATTGTAATACGACTACTATAGGCG 3'  
 Seq primer: 5' GAATTAACCTCACTAAAGG 3'.

## FEATURES

Location/Qualifiers  
 1..241  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /ab\_host="E. coli XL1-Blue"  
 /clone\_lib="Human fetal heart, Lambda ZAP Express"  
 /note="Vector: Lambda ZAP Express; Site 1: EcoRI; Site 2: XbaI; mRNA was purified from human fetal hearts (9-10 weeks). cDNA was synthesized using a XbaI-Oligo dT

**AUTHORS**  
 Wistow, G., Bernstein, S.L., Wyatt, M.K., Farris, R.N., Behal, A.,  
 Touchman, J.W., Bouffard, G., Smith, D., and Peterson, K.  
**TITLE**  
 Expressed sequence tag analysis of human RPE/choroid for the  
 NEBank Project: Over 6000 non-redundant transcripts, novel genes  
 and splice variants  
**JOURNAL**  
 Mol. Vis. 8 (4), 205-220 (2002)  
**MEDLINE**  
 2103460  
**PUBMED**  
 12107410  
**COMMENT**  
 Contact: Wistow G  
 Section on Molecular Structure and Function  
 National Eye Institute  
 6/331, NIH, Bethesda, MD 20892-2740, USA  
 Tel: 301 402 3452  
 Fax: 301 496 0078  
 Email: graeme@helix.nih.gov  
 Plate: 89 row: c column: 07  
 Seq primer: M13RP1 reverse primer (ABI).  
**FEATURES**  
 source  
 1..531  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="c989c07"  
 /tissue\_type="RPE/choroid"  
 /dev\_stage="Adult"  
 /lab\_host="EMDH10B"  
 /clone\_lib="Human Retinal pigment epithelium/choroid cDNA  
 (UN-normalized, unamplified): cs"  
 /note="Organ: Eye; Vector: pCMVSPORT6; Two different donor  
 eyes (75-80 years old) yielded approximately 600 mg of  
 dissected RPE/choroid tissue. This in turn yielded 340 µg  
 of total RNA and 7 µg of mRNA. A directionally cloned cDNA  
 library in the pCMVSPORT6 vector was constructed at Life  
 Technologies (Rockville, MD; now part of Invitrogen Corp),  
 essentially following the protocols of the SuperScript  
 Plasmid System (Invitrogen Corp).  
 <http://www.invitrogen.com/>). The library code  
 designation was cs. For this library, cDNA inserts were  
 cloned into the NotI/MluI sites of the vector. EST  
 analysis was performed on the unamplified library at the  
 NIH Intramural Sequencing Center (NISC)."  
 149 a 87 c 107 g 188 t  
 BASE COUNT  
 ORIGIN  
 Query Match 3.0%; Score 85; DB 14; Length 531;  
 Best Local Similarity 100.0%; Pred. No. 2.2e-31;  
 Matches 85; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2195 AGAGAGGTGAAGAGGAGAACTGGTGTCCGAGGTGCCATTGGATCAAAAGGAGAACTCG 2254  
 |||||||  
 DB 447 AGAGAGGTGAAGAGGAGAACTGGTGTCCGAGGTGCCATTGGATCAAAAGGAGAACTCG 506  
 |||||||  
 QY 2255 GGGTGGATGGCTTCGATGGGGCCCGC 2279  
 |||||||  
 DB 507 GGGTGGATGGCTTCGATGGGGCCCGC 531  
 |||||||  
**RESULT 21**  
**LOCUS** R58412  
**DEFINITION** G3577 Fetal heart Homo sapiens cDNA clone G3577 5' end, mRNA  
 sequence.  
**ACCESSION** R58412  
**VERSION** R58412.1 GI:828470  
**KEYWORDS** EST.  
**SOURCE** Homo sapiens (human)  
**ORGANISM** Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
**REFERENCE** 1 (bases 1 to 161)  
**AUTHORS** Hwang, D.M., Fung, Y.W., Wang, R.X., Laurensen, C.M., Ng, S.H., Lam  
 W.Y., Tsui, K.W., Fung, K.P., Waye, M., Lee, C.Y. and Liew, C.C.  
 Analysis of expressed sequence tags from a fetal human heart cDNA

**JOURNAL**  
 Genomics 30 (2), 293-298 (1995)  
**MEDLINE**  
 96163883  
**PUBMED**  
 8586430  
**COMMENT**  
 Contact: Liew CC  
 Brigham and Women's Hospital  
 Harvard Medical School  
 75 Francis St. Boston, MA 02115, USA  
 Tel: 6177328915  
 Fax: 6179750995  
 Email: cliew@rics.bwh.harvard.edu  
 Seq primer: GGGCGACGACTCCCTGGAGCC.  
**FEATURES**  
 Location/Qualifiers  
 1..161  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="G3577"  
 /lab\_host="E. coli Y1090"  
 /note="Vector: Lambda gt22; Site 1: NotI; Site 2: SalI;  
 mRNA was purified from human fetal hearts (10-12 weeks).  
 cDNA was constructed using a NotI-Oligo dT adaptor-primer.  
 SalI adaptors were ligated, followed by digestion with  
 NotI, for direction cloning into predigested lambda gt22.  
 Method is described in J. Mol. Cell. Cardiol. (1994) 26,  
 1329-1333"  
 51 a 29 c 50 g 31 t  
 BASE COUNT  
 ORIGIN  
 Query Match 2.6%; Score 74; DB 14; Length 161;  
 Best Local Similarity 100.0%; Pred. No. 6.5e-26;  
 Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1854 GGAATCCAGGATTCCTGGAAACCGAGGATTAATGCCCAAAAGCGAGAAATGGGCTT 1863  
 |||||||  
 DB 52 GGAATCCAGGATTCCTGGAAACCGAGGATTAATGCCCAAAAGCGAGAAATGGGCTT 111  
 |||||||  
 QY 1864 CCAGACAGCAAGG 1877  
 |||||||  
 DB 112 CCAGACAGCAAGG 125  
 |||||||  
**RESULT 22**  
**LOCUS** BE701852  
**DEFINITION** CM1-NN1006-30050C-231-gc1 NN1006 Homo sapiens cDNA, mRNA sequence.  
**ACCESSION** BE701852  
**VERSION** BE701852.1 GI:10089594  
**KEYWORDS** EST.  
**SOURCE** Homo sapiens (human)  
**ORGANISM** Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
**REFERENCE** 1 (bases 1 to 330)  
**AUTHORS** Dias Neto, E., Garcia Correa, R., Vetrovski-Almeida, S., Briones, M.R.,  
 Nagai, M.A., da Silva, M. Jr., Zago, M.A., Bordin, S., Costa, F.F.,  
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,  
 Brunsteit, A., de Oliveira, P.S., Eucher, P., Jongeneel, C.V., O'Hare,  
 M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and  
 Simpson, A.J.  
 Shotgun sequencing of the human transcriptome with ORF expressed  
 sequence tags  
**TITLE** Shotgun sequencing of the human transcriptome with ORF expressed  
 sequence tags  
**JOURNAL** Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000);  
**MEDLINE** 20202663  
**PUBMED** 10737800  
**COMMENT**  
 Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
 Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001

```

Best Local Similarity 100.0%; Pred. No. 4.1e-42;
Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2132 AAAATGAAGACAGGAAATTCAGGGCAACAGGGAATTCAGGGCAATCATGTCGAAAAG 2:91
Db 1 AAAATGAAGACAGGAAATTCAGGGCAACAGGGAATTCAGGGCAATCATGTCGAAAAG 60

QY 2192 GAGAGAGGTGAAAGGGAGAACTCGTGTCCGAGGTGCCATTGG 2237
Db 61 GAGAGAGGTGAAAGGGAGAACTCGTGTCCGAGGTGCCATTGG 106

RESULT 18
AA451616 485 bp mRNA linear EST 05-JUN-1997
LOCUS 2x43b04.t1 Soares total fetus Nb2HF8 9w Homo sapiens cDNA clone
DEFINITION IMAGE:789199 5' similar to TR:G2923352 G292352 COLLAGEN CHAIN RH ;,
mRNA sequence.
ACCESSION AA451616
VERSION AA451616.1 GI:2165285
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 485)
AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B.,
Schattenberg, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wylie,
T., Waterston, R. and Wilson, R.
TITLE WashU-Merck EST Project 1997
JOURNAL Unpublished
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 422.
FEATURES
source
1..485
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:5986816"
/db_xref="taxon:9606"
/clone="IMAGE:789199"
/dev_stage="8-9 weeks"
/lab_host="DH10B"
/clone_lib="Soares total_fetus Nb2HF8 9w"
/notes="Vector: pRTT3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from pooled 8-9 week
(total) fetus material with a Not I - oligo(dT) primer (5'
TGTTACCAATCTGAGTGGAGCGCGCTTAATTTTTTTTTTTT 3').
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pRTT3 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaudo."
BASE COUNT 160 a 86 c 91 g 148 t
ORIGIN
Query Match 3.6%; Score 104; DB 9; Length 485;
Best Local Similarity 100.0%; Pred. No. 4.4e-41;
Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2397 AGCCAGCTACAGTCTTACTTCAGAGTGGAGAAATTAGAAATTGATCATTCCTGTC 2456
Db 64 AGCCAGCTACAGTCTTACTTCAGAGTGGAGAAATTAGAAATTGATCATTCCTGTC 123

```

```

QY 2457 CCAACATGGTCCCGGGTATTCCTGGGCCACCTGGTCCGATAG 2503
Db 124 CCAACATGGTCCCGGGTATTCCTGGGCCACCTGGTCCGATAG 167

RESULT 19
N85186 387 bp mRNA linear EST 01-APR-1996
LOCUS J2812F Human fetal heart, Lambda ZAP Express Homo sapiens cDNA
DEFINITION clone J2812 5' similar to REPETITIVE ELEMENT, mRNA sequence.
ACCESSION N85186
VERSION N85186.1 GI:1260811
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 387)
AUTHORS Liew, C.-C.
TITLE CDNAs from fetal heart (1996)
JOURNAL Unpublished
COMMENT Contact: Liew CC
Brigham and Women's Hospital
Harvard Medical School
75 Francis St. Boston, MA 02115, USA
Tel: 6177328915
Fax: 6179750995
Email: cliew@rics.bwh.harvard.edu
Seq primer: GAAATTAACCTCTACTAAAGG.
FEATURES
source
1..387
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="J2812"
/lab_host="E. coli XL1-Blue"
/clone_lib="Human fetal heart, Lambda ZAP Express"
/notes="Vector: Lambda ZAP Express; Site 1: EcoRI; Site 2:
XhoI; mRNA was purified from human fetal hearts (8-10
weeks). cDNA was synthesized using a XhoI-Oligo dT
adaptor-primer. EcoRI adaptors were ligated, followed by
digestion with XhoI, for directional cloning into
predigested lambda ZAP express."
BASE COUNT 123 a 73 c 126 g 65 t
ORIGIN
Query Match 3.0%; Score 85; DB 14; Length 387;
Best Local Similarity 100.0%; Pred. No. 2e-31;
Matches 85; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2:33 AAATGAGACAGGGAATTCAGGGCAACAGGGAATTCAGGCCATCATGTCGAAAAGG 2192
Db 122 AAATGAGACAGGGAATTCAGGGCAACAGGGAATTCAGGCCATCATGTCGAAAAGG 181

QY 2193 AGAGAGAGGTGAAAAGGGAGAACCT 2217
Db 182 AGAGAGAGGTGAAAAGGGAGAACCT 206

RESULT 20
CA397287 531 bp mRNA linear EST 06-NOV-2002
LOCUS cs89c07.y1 Human Retinal pigment epithelium/choroid cDNA
DEFINITION 5', mRNA sequence.
ACCESSION CA397287
VERSION CA397287.1 GI:24734456
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 531)

```

Best Local Similarity 100.0%; Pred. No. 7.2e-54;  
Matches 129; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2153 CAGGCAACAGGGAATTCAAGCCCATCATGTGCAAAAGGAGAGAGAGTGAAGAGGAG 2212  
|||||  
Db 22 CAGGCAACAGGGAATTCAAGCCCATCATGTGCAAAAGGAGAGAGAGTGAAGAGGAG 81  
|||||  
QY 2213 AACCTGGTCCGAGGTCCTCATTCGATCAAAAGGAGAGATCTGGGTGGATGGCTTGATGG 2272  
|||||  
Db 82 AACCTGGTCCGAGGTCCTCATTCGATCAAAAGGAGAGATCTGGGTGGATGGCTTGATGG 141  
|||||  
QY 2273 GGCCCGCAG 2281  
|||||  
Db 142 GGCCCGCAG 150  
|||||

RESULT 16  
CB852087/c  
LOCUS  
DEFINITION  
CB852087  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 792)  
Bonaldo, M.F., Lennon, G. and Soares, M.B.  
Normalization and subtraction: two approaches to facilitate gene  
discovery  
Genome Res. 6 (9), 791-806 (1996)  
97044477  
8889548  
Contact: McCray, PB  
University of Iowa  
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA  
Tel: 319 356 4866  
Fax: 319 356 7171  
Email: paul-mccray@iowa.edu

Tissue Procurement: Dr. M. J. Welsh, University of Iowa  
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Researchers may obtain clones from Research  
Genetics (www.resgen.com) or from Open Biosystems  
(www.openbiosystems.com).  
Seq primer: M13 FORWARD  
POLYA=No. Location/Qualifiers  
1..792  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="UI-CF-FNO-aep-m-17-0-UI"  
/tissue\_type="Human Lung Epithelial cells"  
/lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
/clone\_lib="UI-CF-FNO"  
/note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a  
modified polylinker; Site 1: EcoRI; Site 2: Not I;  
UI-CF-FNO is a subtracted cDNA library derived from two  
normalized Human lung epithelial cell libraries (EN1 and  
DUL) The library was subtracted according to according to  
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,  
1996. For additional information, contact:  
bento-soares@iowa.edu  
TAG SEQ=None found"

BASE COUNT 245 a 174 c 121 g 252 t  
ORIGIN  
Query Match 3.9%; Score 112; DB 14; Length 792;

Best Local Similarity 100.0%; Pred. No. 4.2e-45;  
Matches 112; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1532 AGGTGATCGTGGACTTCTCTGGTTTCTCTGGCTTATGGCATGCCAGGATCAAGGGGTG 1591  
|||||  
Db 300 AGGTGATCGTGGACTTCTCTGGTTTCTCTGGCTTATGGCATGCCAGGATCAAGGGGTG 241  
|||||  
QY 1592 AAATGGGTGCAAGGAGACAAAGGATCACCTGGATTTTATGGCAAAGGG 1643  
|||||  
Db 240 AAATGGGTGCAAGGAGACAAAGGATCACCTGGATTTTATGGCAAAGGG 189  
|||||

RESULT 17  
AA329319  
LOCUS  
DEFINITION  
AA329319  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 452)  
Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A., Buit  
C.J., Lee, N.H., Kirkness, E.F., Weinstein, K.G., Gocayne, J.D., White  
O., Sutton, G., Blake, J.A., Brandon, R.C., Man-Wai, C., Clayton, R.A.,  
Cline, T.R., Cotton, M.D., Earle-Hughes, J., Fine, L.D., Fitzgerald,  
J.M., Fitzhugh, W.M., Fritchman, J.L., Geoghagen, N.S., Glodak, A.,  
Gnehm, C.L., Hanna, M.C., Hedblom, E., Hinkle, P.S., Jr., Kelley, J.M.,  
Kelley, J.C., Liu, L.-I., Marmaros, S.M., Merrick, J.M.,  
Moreno-Balances, R.F., McDonald, S.A., Nguyen, D.T., Pelligrino, S.M.,  
Phillips, C.A., Ryder, S.E., Scott, J.L., Sauder, D.M., Shirley, R.,  
Small, K.V., Spriggs, T.A., Utterback, T.R., Weidman, J.F., Li, Y.,  
Bedratik, D.P., Cao, J., Cepeda, M.A., Coleman, T.A., Collins, E.J.,  
Dimke, D., Feng, D.-F., Fertie, A., Fischer, C., Hastings, G.A., He, W.W.,  
Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K., Kosak, D.L.,  
Karsch, C., Hungjun, C., Li, H., Weissner, P.S., Olsen, H., Raymond, L.,  
Wei, Y.F., Wang, J., Xu, C., Yu, G.L., Ruben, S.M., Dillon, P.J., Fannor  
M.R., Rosen, C.A., Haseeltine, W.A., Fields, C., Fraser, C.M. and  
Venter, J.C.  
Initial assessment of human gene diversity and expression patterns  
based upon 83 million nucleotides of cDNA sequence  
Nature 377 (6547 Suppl. 3-174 (1995))  
96026280  
7566098  
Other ESTs: THC180973  
Contact: Kerlavage, AR  
Bioinformatics  
The Institute for Genomic Research  
9712 Medical Center Drive, Rockville, MD 20850 USA  
Tel: 3018699056  
Fax: 3018699423  
Email: arkerlav@tigr.org  
For clone availability, additional sequence and expression  
information related to this EST, please check the TIGR Human Gene  
Index (<http://www.tigr.org/tdb/hg/hgi.html>)  
Seq primer: M13 Reverse.  
Location/Qualifiers  
1..452  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="ATCC (inhost):130967"  
/db\_xref="taxon:9606"  
/dev\_stage="embryo, 12 wks"  
/clone\_lib="Embryo, 12 week"  
/note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:  
XhoI"

BASE COUNT 226 a 90 c 132 g 101 t 3 others  
ORIGIN  
Query Match 3.7%; Score 106; DB 9; Length 452;



J.E., Saeed, A.I., Sharov, V., Lee, N.H., Yeatman, T.J., and Quackenbush, J.  
Assessment of gene expression patterns in a model of colon tumor metastasis using a 19,200 element cDNA microarray  
Unpublished  
Contact: John Quackenbush  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 3528  
Fax: 301 838 0208  
Email: johnq@tigr.org  
Plate: 176  
Seq primer: Reverse.  
Location/Qualifiers  
source  
1. 494  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone\_lib="MAGE resequences, MAGE"  
/note="Vector: pBluescriptSKm"  
BASE COUNT 156 a 100 c 137 g 101 t  
ORIGIN

Query Match 5.7%; Score 164; DB 10; Length 494;  
Best Local Similarity 99.2%; Pred. No. 1.1e-71;  
Matches 264; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2133 AAATGGAACACAGGGAATTCAGGGCAACAGGGAATTCAGGCCATCATGTCGAAAGG 2192  
DB 1 AAATGGAACACAGGGAATTCAGGGCAACAGGGAATTCAGGCCATCATGTCGAAAGG 60

QY 2193 AGAGAGAGGTGAAAGGGAGAACCTTGTTCCGAGGTGCCATTTGGATCAAAAGGAGAAATC 2252  
DB 61 AGAGAGAGGTGAAAGGGAGAACCTTGTTCCGAGGTGCCATTTGGATCAAAAGGAGAAATC 120

QY 2253 TGGGGTGGATGCTTGATGGGCGCGCAGTCTTAAGGGCAACCTGGGGATCCAGGTCC 2312  
DB 121 TGGGGTGGATGCTTGATGGGCGCGCAGTCTTAAGGGCAACCTGGGGATCCAGGTCC 180

QY 2313 TCAGGACCCCGAGGTTTGGATGGGAAGCCCGAAGAGAGAGTTTCAGAACAAATTTATTCG 2372  
DB 181 TCAGGACCCCGAGGTTTGGATGGGAAGCCCGAAGAGAGTTTCAGAACAAATTTATTCG 242

QY 2373 ACAGTTTCACAGATGTAATAAGAG 2398  
DB 241 ACAAGTTTCACAGATGTAATAAGAG 266

RESULT 12  
BG699698  
LOCUS  
DEFINITION BG699698 744 bp mRNA linear EST 07-MAY-2001  
602881559F1 NIH\_MGC\_95 Homo sapiens cDNA clone IMAGE:4814199 5';  
mRNA sequence.  
ACCESSION BG699698  
VERSION BG699698.1 GI:13968272  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 744)  
NIH-MGC http://mgi.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-r@mail.nih.gov  
Tissue procurement: Niklos Palkovits, M.D., Ph.D.  
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki Toshiyuki and Piero Carninci (RIKEN)  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov  
Plate: LIA10709 row: e column: 16  
High quality sequence stop: 596.  
Location/Qualifiers  
source  
1. 744  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone\_lib="IMAGE:4814199"  
/tissue\_type="hippocampus"  
/lab\_host="DH10B"  
/clone\_lib="NIH\_MGC\_95"  
/note="Organ: brain; Vector: pBluescriptR (modified pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (gtcgag); Oligo-dT primed using primer 5'-TTTCTTTT-TTTTUN-3', size-selected for average insert size 2.5 kb and normalized to R0.5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIH/NHGRI, National Institutes of Health). Note: this is a NIH\_MGC Library."  
BASE COUNT 0 202'a 206 c 177 g 159 t  
ORIGIN

Query Match 5.7%; Score 163; DB 10; Length 744;  
Best Local Similarity 100.0%; Pred. No. 4.1e-71;  
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCGTCACTATATTACATTTCTCTGCATGTTTGGTGGCTTCTTCAGAAATTCGTG 60  
DB 193 ATGCGTCACTATATTACATTTCTCTGCATGTTTGGTGGCTTCTTCAGAAATTCGTG 252

QY 61 TTAGCTGAAGATGGGGAAGTAAGATCAAGTTGCTGCTACTGCTCGACAGATTTAGTTTC 120  
DB 253 TTAGCTGAAGATGGGGAAGTAAGATCAAGTTGCTGCTACTGCTCGACAGATTTAGTTTC 312

QY 221 ATCTTAGATGCTCTTATAGTTTGGCCCGACAAACTTTGAAA 163  
DB 313 ATCTTAGATGCTCTTATAGTTTGGCCCGACAAACTTTGAAA 355

RESULT 13  
AA095402  
LOCUS  
DEFINITION AA095402 161 bp mRNA linear EST 25-OCT-1996  
14401.Seq.F Human fetal heart, Lambda Zap Express Homo sapiens cDNA  
5', mRNA sequence.  
ACCESSION AA095402  
VERSION AA095402.1 GI:1541019  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 161)  
Liew, C.C.  
cDNAs from fetal heart (1996)  
REFERENCE Unpublished  
AUTHORS Contact: Liew CC  
TITLE Brigham and Women's Hospital  
JOURNAL Harvard Medical School  
COMMENT 75 Francis St. Boston, MA 02115, USA  
Tel: 6177328915  
Fax: 6179750995  
Email: cliew@rics.bwh.harvard.edu  
PCR Primers  
FORWARD: 5' GCCAAGCTCGAATTAACCCCTCACTAAAGG 3'  
BACKWARD: 5' CCAGTAATGTAACTAGACTCACTAGGCG 3'  
Seq primer: 5' GAAATTAACCCCTCACTAAAGG 3'.  
Location/Qualifiers  
source  
1. 161  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"





ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 327)  
AUTHORS Hallier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, N., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.  
TITLE The WashU-Merck EST Project  
JOURNAL Unpublished  
COMMENT Contact: Willson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.edu  
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: T7  
High quality sequence stop: 274.  
FEATURES  
source 1..327  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="GDB:3896080"  
/db\_xref="taxon:9606"  
/clone="IMAGE:277704"  
/sex="male"  
/tissue\_type="multiple sclerosis lesions"  
/dev\_stage="Age 46"  
/lab\_host="DH10B (ampicillin resistant)"  
/clone\_lib="Soares multiple sclerosis 2NBHMSR"  
/note="Vector: p77T3D (Pharmacia) with a modified polylinker V type: phagemid. Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(3T) primer [5].  
IGTACCACCTGAGTGGAGCGCGCGCATTTTCTTTTCTTTTCTTTT 3'  
double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified p77T3 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M. Fatima Bonaldo. RNA from 4 multiple sclerosis lesions from one patient was kindly provided by Dr. Kevin G. Becker (NINDS/NIH)."  
BASE COUNT 82 a 80 c 100 g 64 t 1 others  
ORIGIN  
Query Match 8.6%; Score 245; DB 14; Length 327;  
Best Local Similarity 100.0%; Pred. No. 5.1e-113;  
Matches 245; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2501 GCCCAGAGGTCCTCCAGAGGATTAACCTGGTTTCCAGGAAGAGATGGTGTCTCTGGATTAG 2560  
DB 74 GCCCAGAGGTCCTCCAGAGGATTAACCTGGTTTCCAGGAAGAGATGGTGTCTCTGGATTAG 133  
QY 2561 TGGGTGTCCTCCAGAGTCCTCCAGAGGATTAACCTGGTTTCCAGGAAGAGATGGTGTCTCTGGATTAG 2620  
DB 134 TGGGTGTCCTCCAGAGTCCTCCAGAGGATTAACCTGGTTTCCAGGAAGAGATGGTGTCTCTGGATTAG 193  
QY 2621 AAAAAGGGAGCCAAAGGTTTGGGTATCTCTGGAGAACAAAGTCTCTCTGGTCCCCCAGGTC 2680  
DB 194 AAAAAGGGAGCCAAAGGTTTGGGTATCTCTGGAGAACAAAGTCTCTCTGGTCCCCCAGGTC 253  
QY 2681 CAGAGGCGCTCTCTGGAGTAAGCAAGAGTCTCTCCAGGAGACCCAGGTCCTCTGGCA 2740  
DB 254 CAGAGGCGCTCTCTGGAGTAAGCAAGAGTCTCTCCAGGAGACCCAGGTCCTCTGGCA 313  
QY 2741 AAGAT 2745  
DB 314 AAGAT 318

RESULT 8  
LOCUS AA227783 233 bp mRNA linear EST 06-AUG-1997  
DEFINITION z57h04.r1 Soares.NHMPU.S1 Homo sapiens cDNA clone IMAGE:667543 5' similar to SW:CAIF\_HUMAN Q07092 COLLAGEN ALPHA 1(XVI) CHAIN  
PRECUSOR: 1; mRNA sequence.  
ACCESSION AA227783  
VERSION AA227783  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 233)  
AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisler, G., Jost, S., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wyllie, T., Waterston, R. and Wilson, R.  
TITLE WashU-Merck EST Project 1997  
JOURNAL Unpublished  
COMMENT Contact: Willson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.edu  
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 1441 Std Error: 0.00  
Seq primer: -28ml3 rev2 ET from Amer-Sham  
High quality sequence stop: 217.  
FEATURES  
source 1..233  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="GDB:5561510"  
/db\_xref="taxon:9606"  
/clone="IMAGE:667543"  
/tissue\_type="pooled human melanocyte, fetal heart, and pregnant uterus"  
/lab\_host="DH10B"  
/clone\_lib="Soares NHMPU S1"  
/note="Organ: mixed (see below); Vector: p77T3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (melanocyte 2NBHM, pregnant uterus NBHPU, and fetal heart NBH19W) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 26C232-265223, 340488-345479, and 484488-489479."  
BASE COUNT 65 a 63 c 67 g 38 t  
ORIGIN  
Query Match 8.1%; Score 233; DB 9; Length 233;  
Best Local Similarity 100.0%; Pred. No. 6.1e-107;  
Matches 233; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2599 GGCCTACCCAGGAAGAAATGGGAAAAAGGGAGCCAGGGTTTGGGTATCTCTGGAGACAA 2658  
DB 1 GGCCTACCCAGGAAGAAATGGGAAAAAGGGAGCCAGGGTTTGGGTATCTCTGGAGACAA 60  
QY 2659 GGTCTCTCTGTCTCCCGAGGTCCTCCAGAGGCGCTCTCTGGAATACCAAGAGAGTCTCTCCA 2718  
DB 61 GGTCTCTCTGTCTCCCGAGGTCCTCCAGAGGCGCTCTCTGGAATACCAAGAGAGTCTCTCCA 120  
QY 2719 GGAGACCCAGGTCCTCTCCCGAGGTCCTCCAGAGGCGCTCTCTGGAATACCAAGAGAGTCTCTCCA 2778  
DB 121 GGAGACCCAGGTCCTCTCCCGAGGTCCTCCAGAGGCGCTCTCTGGAATACCAAGAGAGTCTCTCCA 160  
QY 2779 CCAGGCGCGCGCGAGGTCCTCCAGAGGCGCTCTCTGGAATACCAAGAGAGTCTCTCCA 283;

Tel: 301 402 3452  
 Fax: 301 496 0078  
 Email: graeme@helix.nih.gov  
 Plate: 68 row: a column: 05  
 Seq primer: M13RP1 reverse primer (AB1).  
 Location/Qualifiers  
 1..527  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="cs68a05"  
 /tissue\_type="RPE/choroid"  
 /dev\_stage="Adult"  
 /lab\_host="EMDH108"  
 /clone\_lib="Human Retinal pigment epithelium/choroid cDNA  
 (Un-normalized, unamplified): cs"  
 /note="Organ: Eye; Vector: pCMVSPORT6; Two different donor  
 eyes (75-80 years old) yielded approximately 600 mg cf  
 dissected RPE/choroid tissue. This in turn yielded 340 ug  
 of total RNA and 7 ug of mRNA. A directionally cloned cDNA  
 library in the pCMVSPORT6 vector was constructed at Life  
 Technologies (Rockville, MD; now part of Invitrogen Corp),  
 essentially following the protocols of the SuperScript  
 Plasmid System (Invitrogen Corp  
 <http://www.invitrogen.com/>). The library code  
 designation was cs. For this library, cDNA inserts were  
 cloned into the NotI/MluI sites of the vector. EST  
 analysis was performed on the unamplified library at the  
 NIH Intramural Sequencing Center (NISC)."  
 BASE COUNT 145 a 118 c 126 g 138 t  
 ORIGIN  
 Query Match 13.2%; Score 377; DB 14; Length 527;  
 Best Local Similarity 100.0%; Pred. No. 2.9e-180;  
 Matches 377; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ATGGCTCACTATATACATTTCTCTGCATGGTTTTGGTGGCTTCTTCAGAAATTCGTG 60  
 Db |||||  
 QY 151 ATGGCTCACTATATACATTTCTCTGCATGGTTTTGGTGGCTTCTTCAGAAATTCGTG 210  
 Db |||||  
 QY 61 TTAGCTGAAGTGGGGAAGTAAGATCAAGTTGTCTACTGCTCCGACAGATTTAGTTTTC 120  
 Db |||||  
 QY 121 ATCTAGATGGCTCTTATAGTGTGGCCGACAAACCTTTGAAATAGTGAAAGTGGCTT 180  
 Db |||||  
 QY 271 ATCTTAGATGGCTCTTATAGTGTGGCCGACAAACCTTTGAAATAGTGAAAGTGGCTT 330  
 Db |||||  
 QY 181 GTCAATATCAGAAAACCTTTGACATAGGCGCGAAGTTTATTCAGTTGGAGTGGTTCAA 240  
 Db |||||  
 QY 331 GTCAATATCAGAAAACCTTTGACATAGGCGCGAAGTTTATTCAGTTGGAGTGGTTCAA 390  
 Db |||||  
 QY 241 TATAGTGACTACCTGTCTCGAGATTCCTCTCGGAAGCTATGATTCAGGAGAACATTTG 300  
 Db |||||  
 QY 391 TATAGTGACTACCTGTCTCGAGATTCCTCTCGGAAGCTATGATTCAGGAGAACATTTG 450  
 Db |||||  
 QY 301 ACGGACAGTGGATCCATCTACTTCTAGGAGAAACACAAAGACAGGAGGCCATC 360  
 Db |||||  
 QY 451 ACGGACAGTGGATCCATCTACTTCTAGGAGAAACACAAAGACAGGAGGCCATC 510  
 Db |||||  
 QY 361 CAGTTTGGCTCGATTA 377  
 Db |||||  
 QY 511 CAGTTTGGCTCGATTA 527  
 Db |||||  
 RESULT 6  
 AI434106/c 276 bp mRNA linear EST 30-MAR-1999  
 LOCUS  
 DEFINITION  
 t131b1.x1 NCI CGAP Gas4 Homo sapiens cDNA clone IMAGE:2132061 3'  
 similar to TR:000339 000339 MARRILIN-2 ;, mRNA sequence.  
 ACCESSION  
 AI434106  
 VERSION  
 AI434106.1 GI:4293857  
 KEYWORDS  
 EST.

SOURCE  
 ORGANISM  
 Homo sapiens (human)  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE  
 1 (bases 1 to 276)  
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 AUTHORS  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL  
 Unpublished  
 COMMENT  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-r@mail.nih.gov  
 Tissue Procurement: Christopher Moshaluk, M.D., Ph.D., Michael R.  
 Emmert-Buck, M.D., Ph.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LEND at:  
 www-bio.llnl.gov/bbrp/image/image.html  
 Insert Length: 893 Std Error: 0.00  
 Seq primer: -40UP from Gibco.  
 FEATURES  
 1..276  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2132061"  
 /tissue\_type="poorly differentiated adenocarcinoma with  
 signet ring cell features"  
 /lab\_host="DH108"  
 /clone\_lib="NCI-CGAP Gas4"  
 /note="Organ: stomach; Vector: pCMV-SPORT6; Site 1: SalI;  
 Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
 Average insert size 1.69 kb. Life Technologies catalog #:  
 '11549-011"  
 BASE COUNT 82 a 61 c 39 g 94 t  
 ORIGIN  
 Query Match 9.6%; Score 276; DB 9; Length 276;  
 Best Local Similarity 100.0%; Pred. No. 7.6e-129;  
 Matches 276; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 477 AGATAGTAAGATAACATTATTGCTATTGGTGGTTCAGAAACAGAGATGCGGAAC 536  
 Db |||||  
 QY 276 AGATAGTAAGATAACATTATTGCTATTGGTGGTTCAGAAACAGAGATGCGGAAC 217  
 Db |||||  
 QY 537 TAGAGCTATTGCCAACAAAGCCTTCGCTACTTATGCTGTTTATGTGGAAGACTATATTGC 596  
 Db |||||  
 QY 216 TAGAGCTATTGCCAACAAAGCCTTCGCTACTTATGCTGTTTATGTTGGAAGACTATATTGC 157  
 Db |||||  
 QY 597 ATATCCAAATTAAGGGAAGTATGAAGCAGAAACCTTGTGAAGAACTCTGCTGCCAAC 656  
 Db |||||  
 QY 156 AATATCCAAATTAAGGGAAGTATGAAGCAGAAACCTTGTGAAGAACTCTGCTGCCAAC 97  
 Db |||||  
 QY 657 ACCAATTCAGTGGCAGCTCGTGATGAAGGGGATTTCATATTCCTTTAGGTTTAGATGT 716  
 Db |||||  
 QY 96 ACGAATTCAGTGGCAGCTCGTGATGAAGGGGATTTCATATTCCTTTAGGTTTAGATGT 37  
 Db |||||  
 QY 717 AATAAAGAGTAAAGAAAGAAATACAGCTTTCACC 752  
 Db |||||  
 QY 36 AATAAAGAGTAAAGAAAGAAATACAGCTTTCACC 1  
 Db |||||  
 RESULT 7  
 N46880  
 LOCUS  
 DEFINITION  
 Y55a01.1 Soares multiple sclerosis 2NBHSP Homo sapiens cDNA  
 clone IMAGE:277704 5' similar to SW:CA1H MOUSE P39061 COLLAGEN  
 ALP5A1(XVII); CHAIN PRECURSOR. ;, mRNA sequence.  
 ACCESSION  
 N46880  
 VERSION  
 N46880.1 GI:1188C46  
 KEYWORDS  
 EST.  
 Homo sapiens (human);  
 SCJCE

```
Db 188 CGAACCAGGAACATGCAATGATGATTAATGGGTAGTCCCGGTTTCAAGGAGAA 247
QY 1756 GCAGGATCCCTCGTCTCCGGGCGAGTGAACACGGGAGAGCTGGAATCCCGAG 1815
Db 248 GCAGGATCCCTCGTCTCCGGGCGAGTGAACACGGGAGAGCTGGAATCCCGAG 307
QY 1816 TTTCCTGGAACCGAGGATTAATGGGCGAAGAGGAGAAATTTGGGCTCCAGGACAGCAA 1875
Db 308 TTTCCTGGAACCGAGGATTAATGGGCGAAGAGGAGAAATTTGGGCTCCAGGACAGCAA 367
QY 1876 GGAAAAAAGAGGCCCGCAGGATGCTGTTTAAATGGGAAGCAATGCTCAGAGCCAG 1935
Db 368 GGAAAAAAGAGGCCCGCAGGATGCTGTTTAAATGGGAAGCAATGCTCAGAGCCAG 427
QY 1936 CCTCGAACACCGGATCTAAGGGAAGCAAGGTGAACCTGGAATTCAGGGATGCTCGG 1995
Db 428 CCTCGAACACCGGATCTAAGGGAAGCAAGGTGAACCTGGAATTCAGGGATGCTCGG 487
QY 1996 GCTTCTGGGCTCAAGGAGAACACAGGAGCAACGGGTTTCCCGAGAGAACAGGATACATG 2055
Db 488 GCTTCTGGGCTCAAGGAGAACACAGGAGCAACGGGTTTCCCGAGAGAACAGGATACATG 547
QY 2056 GGTTCACCGGATTCAGGAAAGAGGAGGAGCAAGGAAATCAAGTGAAGAGTATT 2115
Db 548 GGTTCACCGGATTCAGGAAAGAGGAGGAGCAAGGAAATCAAGTGAAGAGTATT 607
QY 2116 CAGGCTCAAAAGGGAGAGAAATGGAAGACAGGGAATTCAGGGCAACAGGGAATTCAGGCG 2175
Db 608 CAGGCTCAAAAGGGAGAGAAATGGAAGACAGGGAATTCAGGGCAACAGGGAATTCAGGCG 667
QY 2176 CATCATGGTCAAAAGGAGAGAGAGGTGAAGAGGAGAACCTGGTTCGAGGTGCAT 2235
Db 668 CATCATGGTCAAAAGGAGAGAGAGGTGAAGAGGAGAACCTGGTTCGAGGTGCAT 727
QY 2236 GGATCAAAAGGAGATCTGGG 2257
Db 728 GGATCAAAAGGAGATCTGGG 749

RESULT 4
BX431407 909 bp mRNA linear EST 15-MAY-2003
LOCUS BX431407 Homo sapiens PLACENTA Homo sapiens cDNA clone CS0DE009YK07
DEFINITION 5-PRIME, mRNA sequence.
ACCESSION BX431407
VERSION BX431407.1 GI:30773057
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Li.W.B., Gruber,C., Jesse,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 3377.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
CGI-bin/cluster.cgi?seq=CS0BAG0132B02_CS01162_1&cluster=3377.r.
Contact : Feng liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0BAG0132B02_CS01162_1.
Location/Qualifiers
1..909
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DE009YK07"
```

```
/tissue type="PLACENTA"
/clone lib="Homo sapiens PLACENTA"
/notes=vector: PCMVSPORT 6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched,
double-strand cDNA was digested with Not I and cloned into
the Not I and EcoRV sites of the PCMVSPORT 6 vector.
Library was not normalized.
BASE COUNT 251 a 201 c 205 g 249 t 3 others
ORIGIN
Query Match 15.2%; Score 436; DB 13; Length 909;
Best Local Similarity 100.0%; Pred. No. 2.8e-210;
Matches 436; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2430 AATTAGAAATTTGATCATTTGCTCCCAACATGGCTCCCGGGTATTCTTGGGCGACC 2489
Db 35 AATTAGAAATTTGATCATTTGCTCCCAACATGGCTCCCGGGTATTCTTGGGCGACC 94
QY 2490 TGTCCGATAGGCCAGAGGGTCCAGAGGATTACCTGGTTTGGCAGGAAGAGATGGTGT 2549
Db 95 TGTCCGATAGGCCAGAGGGTCCAGAGGATTACCTGGTTTGGCAGGAAGAGATGGTGT 154
QY 2550 TCTTGGGATTAGTGGGTGCTCCCTGGAGCTCCAGGTGTCAGAGGATTAAGAGGCGCTACCGAG 2609
Db 155 TCTTGGGATTAGTGGGTGCTCCCTGGAGCTCCAGGTGTCAGAGGATTAAGAGGCGCTACCGAG 214
QY 2610 AAGAAATGGGGAAGAGGAGGAGGAGGTTTGGGTATCTTGGAGAACAGGTCTTCTGG 2669
Db 215 AAGAAATGGGGAAGAGGAGGAGGAGGTTTGGGTATCTTGGAGAACAGGTCTTCTGG 274
QY 2670 TCCCGGAGTCCAGAGGGCTCTCTGGAAATAGCAAGAGGTTCTCCAGGAGACCGAG 2729
Db 275 TCCCGGAGTCCAGAGGGCTCTCTGGAAATAGCAAGAGGTTCTCCAGGAGACCGAG 334
QY 2730 TCTCCCTGGCAAGATGGAGACCATGGAAACCTGGAAATCCAGGGCAACAGGCCCGCC 2789
Db 335 TCTCCCTGGCAAGATGGAGACCATGGAAACCTGGAAATCCAGGGCAACAGGCCCGCC 334
QY 2790 AGGCATCTGGACCCATCATCTATTTTAGTGTAAATGGCAGGAAGAGATCCGTTCAAGAA 2849
Db 395 AGGCATCTGGACCCATCATCTATTTTAGTGTAAATGGCAGGAAGAGATCCGTTCAAGAA 454
QY 2850 AGGACCAAACTATTAG 2865
Db 455 AGGACCAAACTATTAG 470

RESULT 5
CA395709 527 bp mRNA linear EST C6-NOV-2002
LOCUS CA395709
DEFINITION CS88a05.y1 Human Retinal pigment epithelium/choroid cDNA
5', mRNA sequence.
ACCESSION CA395709
VERSION CA395709.1 GI:24731429
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 527)
AUTHORS Wistow,G., Bernstein,S.L., Wyatt,M.K., Farris,R.N., Behai,A.,
Touchman,J.W., Bouffard,G., Smith,D. and Peterson,K.
TITLE Expressed sequence tag analysis of human RPE/choroid for the
NEI/RP Project: Over 6000 non-redundant transcripts, novel genes
and splice variants
Mol. Vis. 8 (4), 205-220 (2002)
JOURNAL Mol. Vis. 8 (4), 205-220 (2002)
MEDLINE 22103460
PUBMED 12107410
COMMENT Contact: Wistow G
Section on Molecular Structure and Function
National Eye Institute
6/331, NIH, Bethesda, MD 20892-2740, USA
```

## COMMENT

Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: Dr. James R. Lupski  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LMNL at:  
<http://image.llnl.gov>  
 Plate: LLM13569 row: h column: 01  
 High quality sequence stop: 693.  
 Location/Qualifiers  
 1. 929  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:6182832"  
 /sex="male"  
 /tissue\_type="dorsal root ganglia"  
 /dev\_stage="adult, 36 yr"  
 /lab\_host="DH10B"  
 /clone\_lib="Lupski dorsal root ganglion"  
 /note="Vector: pCMV-SPORT6 (Life Technologies); Site 1:  
 NotI; Site 2: SalI; cDNA made by oligo-dT priming.  
 Directionally cloned using the following adaptors:  
 5'-TCGACCCACGGCTCGG-3' and  
 5'-GACTAGTTCTAGATCGAGCGCGCCCT(15)-3'. Size selected >  
 1 kb for average insert length 1.7 kb. This is a primary  
 library, non-amplified. Library constructed by Life  
 Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor  
 College of Medicine) and is available through Life  
 Technologies."

BASE COUNT 254 a 209 c 256 g 210 t

## FEATURES

source

## RESULT 3

EX458795

LOCUS

DEFINITION

EX458795

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Db 481 GGAGCCAAAGGCTTTGGGTATCTCTGGAGAAAGAGTCTCTCTGGTCCCTCCAGAGTCCAGAG 540  
 QY 2686 GGCCTCTCTGGATAGCAAGAGAGTCTCTCCAGAGACCCAGGCTCTCTGGCAAGAT 2745  
 Db 541 GGCCTCTCTGGATAGCAAGAGAGTCTCTCCAGAGACCCAGGCTCTCTGGCAAGAT 600  
 QY 2746 GGAGACCATGGAAACCTCGAATCCAAAGGCAACCCAGGCCCCCAGGAGTCTGGACCCA 2805  
 Db 601 GGAGACCATGGAAACCTCGAATCCAAAGGCAACCCAGGCCCCCAGGAGTCTGGACCCA 660  
 QY 2806 TCATATGTTTATGTTATTCCTCCAGAGAGATTCGTTTCAGAAAAAGGACCAACTATTAG 2865  
 Db 661 TCATATGTTTATGTTATTCCTCCAGAGAGATTCGTTTCAGAAAAAGGACCAACTATTAG 720

EX458795 Homo sapiens PLACENTA mRNA linear EST 22-MAY-2003  
 5-PRIME, mRNA sequence.

EX458795.1 GI:31033003

ES..

Homo sapiens (human)

Homo sapiens

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 1201)

Li W.B., Gruber, C., Jessee, J. and Polayes, D.

Full-length cDNA libraries and normalization

Unpublished

Contact: Genoscope

Genoscope - Centre National de Sequencage

Bp 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, web : www.genoscope.cns.fr

Library was constructed by Life Technologies, a division of

Invitrogen. This sequence belongs to sequence cluster 3377.r For

more information about this cluster, see

<http://www.genoscope.cns.fr/>

CGI-bin/cluster.cgi?seq=CS0DE009AF040P1&cluster=3377.r. Contact :

Peng Liang Email : fliang@lifetech.com URL :

<http://fulllength.invitrogen.com/> Invitrogen Corporation 1600

Faraday Avenue Genoscope sequence ID : CS0DE009AF040P1.

## FEATURES

source

1. 1201

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="CS0DE009YK07"

/tissue\_type="PLACENTA"

/clone\_lib="Homo sapiens PLACENTA"

/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed

with a NotI-oligo(dT) primer. Five prime end enriched,

double-strand cDNA was digested with Not I and cloned into

the Not I and EcoRV sites of the pCMVSPORT 6 vector.

Library was not normalized."

BASE COUNT 350 a 223 c 387 g 212 t

## ORIGIN

Query Match

Best Local Similarity

Matches 682; Conservative

23.8%; Score 682; DB 13; Length 1201;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 682; Conservative

0; Mismatches

0; Indels

0; Gaps

0;

QY 1576 CCAGATCAAAAGGTTGAAATGGGTGCCAAAGAGACAAAGGATCACCTGGATTTATGGC 1635

Db 68 CCAGATCAAAAGGTTGAAATGGGTGCCAAAGAGACAAAGGATCACCTGGATTTATGGC 127

QY 1636 AAAAGGGTCAAAAGGTTGAAATGGGTGCCAAAGAGACAAAGGATCACCTGGATTTATGGC 1695

Db 128 AAAAGGGTCAAAAGGTTGAAATGGGTGCCAAAGAGACAAAGGATCACCTGGATTTATGGC 187

QY 1696 GGAGATCAAAAGGTTGAAATGGGTGCCAAAGAGACAAAGGATCACCTGGATTTATGGC 1755

QY 961 TTATTTCAACAAACAGCGTAAATTAATGGCTCAACAAGTGGTTACCTTTGCTAACCCCTCAA 1020  
DB 1148 TTATTTACAAACACAGCGTAATTAATGGCTTACAAAGTGGTTACCTTTGCTAACCCCTCAA 1207  
QY 1021 GTTAAGACGTTGTTTGTAGTGAAGGCTGGCAACAAATTCGTCTCTTAGTAAACAGACAT 1080  
DB 1208 GTTAAGACGTTGTTTGTAGTGAAGGCTGGCAACAAATTCGTCTCTTAGTAAACAGACAT 1267  
QY 1081 GTGACTTTGTATATTTGATGACCAACAAATTCGAACCAAGCCCTTACATCCAGTTTATGGG 1140  
DB 1268 GTGACTTTGTATATTTGATGACCAACAAATTCGAACCAAGCCCTTACATCCAGTTTATGGG 1327  
QY 1141 ATCTTGATCAATGGGCAACCAAAATTCGAACCAATTCGAACCAAGCCCTTACATCCAG 1200  
DB 1328 ATCTTGATCAATGGGCAACCAAAATTCGAACCAATTCGAACCAAGCCCTTACATCCAG 1387  
QY 1201 TTTGATGTCGAAAGTTGCGAATCTACTGTGACCCAGAACAAACACCGGAGACAGCA 1260  
DB 1388 TTTGATGTCGAAAGTTGCGAATCTACTGTGACCCAGAACAAACACCGGAGACAGCA 1447  
QY 1261 TGTGAGATTCCTGGATTT-----TGCCTTAATGGTCCAGTGATGTAGGTTCAACT 1311  
DB 1448 TGTGAGATTCCTGGATTTAAATGGAGAGTGCCTTAATGGTCCAGTGATGTAGGTTCAACT 1507  
QY 1312 CCAGCTCCCTGTATTTGTCCTCCGGGAAACACAGGACTTCAAGGCCCCAAAGGTGACCC 1371  
DB 1508 CCAGCTCCCTGTATTTGTCCTCCGGGAAACACAGGACTTCAAGGCCCCAAAGGTGACCC 1567  
QY 1372 GGACTGCTCGGAAACCCCTGGCTACCTGGACAACTGGTCAAGATGTAAGCCCTGGATAT 1431  
DB 1568 GGACTGCTCGGAAACCCCTGGCTACCTGGACAACTGGTCAAGATGTAAGCCCTGGATAT 1619  
QY 1432 CAGGGAATTCAGGAGACACCCAGGTTTCCAGGATCTCCAGGAATACAAAGAGTCCAGGA 1491  
DB 1620 CAGGGAATTCAGGAGACACCCAGGTTTCCAGGATCTCCAGGAATACAAAGAGTCCAGGA 1679  
QY 1492 CTACAGGTTACAAAGAGAACACCGGGGAGATGGTGACAGGATGATCGTGACTTCCT 1551  
DB 1680 CTACAGGTTACAAAGAGAACACCGGGGAGATGGTGACAGGATGATCGTGACTTCCT 1739  
QY 1552 GGTTCCTCTGGCTTTTCATGGATGCCAGGATCAAGGCTGAATGGCTGCCAAGAGAGAC 1611  
DB 1740 GGTTCCTCTGGCTTTTCATGGATGCCAGGATCAAGGCTGAATGGCTGCCAAGAGAGAC 1799  
QY 1612 AAAGGATCACCTGGATTTTATGGCAAAAGAGGTGCAAAAGGTGAAAAGGGGAATGCTGGC 1672  
DB 1800 AAAGGATCACCTGGATTTTATGGCAAAAGAGGTGCAAAAGGTGAAAAGGGGAATGCTGGC 1859  
QY 1672 TTCCCTGGCTCCCTGGACCTGCTGGAGAACACAGGACATGGAAGGATGGAATTAATG 1731  
DB 1860 TTCCCTGGCTCCCTGGACCTGCTGGAGAACACAGGACATGGAAGGATGGAATTAATG 1919  
QY 1732 GGTAGTCCCGTTTCAAGGAGAACAGGATCCCTGTGTCTCGGGGAGGATGGAACA 1791  
DB 1920 GGTAGTCCCGTTTCAAGGAGAACAGGATCCCTGTGTCTCGGGGAGGATGGAACA 1979  
QY 1792 CGGGGAGAGCTGGAAATCCAGGATTTCTCGAAGCCAGGATTAATGGGCCAAAGGGA 1851  
DB 1980 CGGGGAGAGCTGGAAATCCAGGATTTCTCGAAGCCAGGATTAATGGGCCAAAGGGA 2039  
QY 1852 GAAATTTGGGCTTCCAGGACAGCAAGGAAAAAAGAGCCCCAGGAGTGCCTGGTTTAATG 1911  
DB 2040 GAAATTTGGGCTTCCAGGACAGCAAGGAAAAAAGAGCCCCAGGAGTGCCTGGTTTAATG 2099  
QY 1912 GGAAGCAATGGCTCACAGGACAGCTGGACACAGGATCTAAGGAGAACAAAGGTGAA 1971  
DB 2100 GGAAGCAATGGCTCACAGGACAGCTGGACACAGGATCTAAGGAGAACAAAGGTGAA 2159  
QY 1972 CTGGAATTCAGGAGTGCCTGGGCTTCTGGCTCAAGGAGAACACAGGAGCAACGGT 2031  
DB 2160 CTGGAATTCAGGAGTGCCTGGGCTTCTGGCTCAAGGAGAACACAGGAGCAACGGT 2219

QY 2032 TCCGAGGAGAACGAGATACATGGTTTACCGGGATTTCAGGAAAAAAGGGGACAAA 2091  
DB 2220 TCCGAGGAGAACGAGATACATGGTTTACCGGGATTTCAGGAAAAAAGGGGACAAA 2279  
QY 2092 GGAATCAAGGTGAAAAAGGTATTACGGGTCAAAAGGGAGAAATGGAAGACAGGGAAT 2151  
DB 2280 GGAATCAAGGTGAAAAAGGTATTACGGGTCAAAAGGGAGAAATGGAAGACAGGGAAT 2339  
QY 2152 CCAGGGCAACAGGAAATTCAGGCGCATCTGGTGCAAAAGGAGAGAGAGTGAAGGGA 2211  
DB 2340 CCAGGGCAACAGGAAATTCAGGCGCATCTGGTGCAAAAGGAGAGAGAGTGAAGGGA 2399  
QY 2212 GAACTGGTCTCCAGAGTGCCTAATGGATCAAAAGGAGAAATCTGGGTTGGATGCTGATG 2271  
DB 2400 GAACTGGTCTCCAGAGTGCCTAATGGATCAAAAGGAGAAATCTGGGTTGGATGCTGATG 2459  
QY 2272 GGGCCCGCAGGTCTCTAAGGGCAACCTGGGATCCAGGTCTCTAGGGACCCCGAGTTG 2331  
DB 2460 GGGCCCGCAGGTCTCTAAGGGCAACCTGGGATCCAGGTCTCTAGGGACCCCGAGTTG 2519  
QY 2332 GATGGGAAGCCCGGAGAGAGTGTTCAGAAACAATTTATTCGACAAAGTTTCACAGATGTA 2391  
DB 2520 GATGGGAAGCCCGGAGAGAGTGTTCAGAAACAATTTATTCGACAAAGTTTCACAGATGTA 2579  
QY 2392 ATAGAGCCCGAGGTACAGTCTTACTTTCAGAGTGGAGAAATTTAGAAATTCGTATTC 2451  
DB 2588 ATAAAGCCCGAGGTACAGTCTTACTTTCAGAGTGGAGAAATTTAGAAATTCGTATTC 2639  
QY 2452 CTGTCCCAACATGCTCCCGGGTATTCCTGGGCCACCTGGTCCGATAGGCCCAAGAGGT 2511  
DB 2640 CTGTCCCAACATGCTCCCGGGTATTCCTGGGCCACCTGGTCCGATAGGCCCAAGAGGT 2699  
QY 2512 CCCAGAGATTACCTGGTTTCCAGGAGAGATGGTCTTCTGATTTAGTGGTGTCCCT 2572  
DB 2720 CCCAGAGATTACCTGGTTTCCAGGAGAGATGGTCTTCTGATTTAGTGGTGTCCCT 2759  
QY 2572 GGAAGTCCAGGTGTCAAGGATTAAGGGCTTACAGGAGAAATGGGAAAGGGAGC 2631  
DB 2760 GGAAGTCCAGGTGTCAAGGATTAAGGGCTTACAGGAGAAATGGGAAAGGGAGC 2819  
QY 2632 CAAGGTTTGGGTATCTCTGGAGAACCAAGGTCTCTCTGGTCCCGCAGGTCCAGAGGGCCCT 2691  
DB 2820 CAAGGTTTGGGTATCTCTGGAGAACCAAGGTCTCTCTGGTCCCGCAGGTCCAGAGGGCCCT 2879  
QY 2692 CTTGGAATTAAGCAAGAGGTCTTCCAGGAGACCCAGGTCTCTCTGGCAAGATGGAGAC 2751  
DB 2880 CTTGGAATTAAGCAAGAGGTCTTCCAGGAGACCCAGGTCTCTCTGGCAAGATGGAGAC 2939  
QY 2752 CATGAAAAACCTGGAATCCAAAGGGCAACCGGGCCCGCAGGCATCTCGACCCCATCACTA 2811  
DB 2940 CATGAAAAACCTGGAATCCAAAGGGCAACCGGGCCCGCAGGCATCTCGACCCCATCACTA 2999  
QY 2812 TGTTTTATGTTAATGGCAGAGAGATCCGTTCAGAAAAAGGACCAAACTATTAG 2865  
DB 3000 TGTTTTATGTTAATGGCAGAGAGATCCGTTCAGAAAAAGGACCAAACTATTAG 3053

RESULT 2  
BQ893743

LOCUS BQ893743 929 bp mRNA linear EST 16-AUG-2002  
DEFINITION AGNCOURT\_8232050 Lupsaki\_dorsal\_root\_ganglion Homo sapiens cDNA  
clone IMAGE:6182832 5', mRNA sequence.

ACCESSION BQ893743  
VERSION BQ893743.1 GI:22285757

KEYWORDS EST  
SOURCE Homo sapiens (human)

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 'bases 1 to 929'

AUTHORS NIH-NSC <http://hgsc.rci.nih.gov/>

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished

```

80 20 0.7 186 9 AI906933
79 20 0.7 246 14 CB241974
80 20 0.7 296 10 BF652344
81 20 0.7 296 9 AW685176
82 20 0.7 296 13 B0557911
83 20 0.7 297 28 B43380
84 20 0.7 323 10 BG583481
85 20 0.7 326 9 AI301744
86 20 0.7 345 28 AZ056334
87 20 0.7 347 28 AQ167717
88 20 0.7 351 28 BH337770
89 20 0.7 353 9 AW356726
90 20 0.7 369 28 AZ039808
91 20 0.7 371 14 CA909116
92 20 0.7 376 29 T122H03Q
93 20 0.7 389 10 BF655157
94 20 0.7 397 9 AA115401
95 20 0.7 406 10 BE284841
96 20 0.7 415 14 C82609
97 20 0.7 415 14 C83465
98 20 0.7 429 14 R23059
99 20 0.7 430 9 AV733646
100 20 0.7 438 9 AV594520

```

# ALIGNMENTS

```

RESULT 1
BC045597 4145 bp mRNA linear HTC 04-MAR-2003
LOCUS Homo sapiens, clone IMAGE:4814199, mRNA.
DEFINITION BC045597
ACCESSION BC045597.1 GI:28374134
VERSION 1 (base 1 to 4145)
KEYWORDS Homo sapiens (human)
SOURCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
ORGANISM Strausberg, R.
Direct Submission
Submitted (31-JAN-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NMGRI) & Shiraki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: http://www-shgc.stanford.edu
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.

```

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>

Series: IRAC Plate: 32 Row: k Column: 15

This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 18780272

This clone has the following problem: retained intron.

## FEATURES

```

source
1..4145
/organism="Homo sapiens"
/mol_type="mRNA"
/db xref="taxon:9606"
/clone="IMAGE:4814199"
/tissue_type="Brain, hippocampus"

```

```

/clone_lib="NIH_MGC_95"
/lab_host="DH10B"
/note="Vector: pBluescript"
BASE COUNT 1275 a 787 c 1017 g 1066 t
ORIGIN
Query Match 54.6%; Score 1564; DB 11; Length 4145;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 2854; Conservative 0; Mismatches 3; Indels 17; Gaps 2;
QY 1 ATGGCTCACTATATACATTTCTGTCATGGTTTGGTGGCTCTCTCAGAAATCTGTG 60
DB -88 AAGGCTCACTATATACATTTCTGTCATGGTTTGGTGGCTCTCTCAGAAATCTGTG 247
QY 61 TTAGCTGAAGATGGGAAGTAAAGTCAAGTTGTCTGCTACTGCTCCGACAGAGATTAGTTTC 120
DB 248 TTAGCTGAAGATGGGAAGTAAAGTCAAGTTGTCTGCTACTGCTCCGACAGAGATTAGTTTC 307
QY 221 ATCTTAGATGCTCTTATAGTTGGTGGCCGAGAACTTTGAAATAGTGAAGAGTGGCTT 180
DB 308 ATCTTAGATGCTCTTATAGTTGGTGGCCGAGAACTTTGAAATAGTGAAGAGTGGCTT 367
QY 381 GTCATATCAAAAAGTCTTGACATAGGCGCGAAGTTTATTCAGTTGGAGTGGTCAA 240
DB 388 GTCATATCAAAAAGTCTTGACATAGGCGCGAAGTTTATTCAGTTGGAGTGGTCAA 427
QY 241 TATAGTACTACCTGTGCTGAGATTCCTCTCGAAGCTATGATTCAGGAGAAATTTG 300
DB 428 TATAGTACTACCTGTGCTGAGATTCCTCTCGAAGCTATGATTCAGGAGAAATTTG 487
QY 301 ACGGAGCAGTGGAAATCATACTCTACTAGGAGGAACACAAAGACAGGGAAGGCCATC 360
DB 488 ACGGAGCAGTGGAAATCATACTCTACTAGGAGGAACACAAAGACAGGGAAGGCCATC 547
QY 361 CAGTTTGGCTCGATTAACCTTTTGGCAAGTCCCTCAGATTTCTGACTAGATAGCAGTG 420
DB 548 CAGTTTGGCTCGATTAACCTTTTGGCAAGTCCCTCAGATTTCTGACTAGATAGCAGTG 607
QY 421 GTACTTACCGATGCAAAATCCCAAGATGACCTCAAGGATGACCTCAAGCAGCAAGAGAT 480
DB 608 GTACTTACCGATGCAAAATCCCAAGATGACCTCAAGGATGACCTCAAGCAGCAAGAGAT 667
QY 481 AGTAAGATAACATTTATTTGCTATTTGGTGGTTCAGAAACAGAGAGTCCGAACTAGA 540
DB 668 AGTAAGATAACATTTATTTGCTATTTGGTGGTTCAGAAACAGAGAGTCCGAACTAGA 727
QY 541 GCTATTGCCAACAGGCTTCGCTACTTATGCTGTTTATGTGGAGGACTATATTCGAATA 600
DB 728 GCTATTGCCAACAGGCTTCGCTACTTATGCTGTTTATGTGGAGGACTATATTCGAATA 787
QY 601 TCCAAAATAGGGAAGTATGAAAGCAGAACTTTGTGAAGAAATCTGTGTGCCAACAGCA 660
DB 788 TCCAAAATAGGGAAGTATGAAAGCAGAACTTTGTGAAGAAATCTGTGTGCCAACAGCA 847
QY 661 ATTCAGTGGCAGCTCGTGATGAAGGGAATTCAGATTTCTTTAGGTTAGATGTAAT 720
DB 948 ATTCAGTGGCAGCTCGTGATGAAGGGAATTCAGATTTCTTTAGGTTAGATGTAAT 907
QY 721 AAAAAGGTTAAGAAAAGATACAGCTTTCACCAAAAAGATATAAGATGAAGTAACA 780
DB 908 AAAAAGGTTAAGAAAAGATACAGCTTTCACCAAAAAGATATAAGATGAAGTAACA 967
QY 781 TCAAAGTGAATTTATCAGAACTCAGAACTCAAGCAATGTTTCCAGAGGCTCTTCCTCCATCA 840
DB 968 TCAAAGTGAATTTATCAGAACTCAGAACTCAAGCAATGTTTCCAGAGGCTCTTCCTCCATCA 1027
QY 841 TATGTAATTTGCTCTACTCAAGATTTAAAGTCAAGAAAATTTGGGATTTATGAGGAATA 900
DB 1028 TATGTAATTTGCTCTACTCAAGATTTAAAGTCAAGAAAATTTGGGATTTATGAGGAATA 1087
QY 901 TTAAGTATTTGAGGAGGCGCAAAATAGCAGTTTACCTTAAATGCTGTGGACAAATCTTA 960
DB 1086 TTAAGTATTTGAGGAGGCGCAAAATAGCAGTTTACCTTAAATGCTGTGGACAAATCTTA 1147

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2003, 22:19:29 ; Search time 5306 Seconds

(without alignments)  
13123.318 Million cell updates/sec

Title: US-09-996-611B-5

Perfect score: 2865  
Sequence: 1 atggtcactattacatt.....gaaaggaccactattag 2865

Scoring table: OLIGO NUC  
Gapop 60.0 , Gapext 60.0

Searched: 22781392 seqs, 12152238056 residues

Word size : 0

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database : EST:

1: em\_estba.\*

2: em\_esthum.\*

3: em\_estin.\*

4: em\_estmu.\*

5: em\_estov.\*

6: em\_estpl.\*

7: em\_estro.\*

8: em\_htc.\*

9: gb\_est1.\*

10: gb\_est2.\*

11: gb\_htc.\*

12: gb\_est3.\*

13: gb\_est4.\*

14: gb\_est5.\*

15: em\_estfun.\*

16: em\_estom.\*

17: em\_ges\_hum.\*

18: em\_ges\_inv.\*

19: em\_ges\_pln.\*

20: em\_ges\_vrt.\*

21: em\_ges\_fun.\*

22: em\_ges\_mam.\*

23: em\_ges\_mus.\*

24: em\_ges\_pro.\*

25: em\_ges\_rtd.\*

26: em\_ges\_phg.\*

27: em\_ges\_vrl.\*

28: gb\_ges1.\*

29: gb\_ges2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1564	54.6	4145	11	BC045597 Homo sapi
2	720	25.1	929	13	BQ93743 AGENCOURT
3	662	23.8	1201	13	BX458795 BX458795
4	436	15.2	909	13	BX431407 BX431407

```

; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5386
; LENGTH: 3609
; TYPE: DNA
; ORGANISM: Aspergillus fumigatus
US-10-128-714-5386

Query Match      0.7%  Score 19;  DB 14;  Length 3609;
Best Local Similarity 100.0%;  Pred. No. 77;
Matches 19;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

QY  1570 GGCATGCCAGGATCAAAGG 1588
      |||||||
Db   3026 GGCATGCCAGGATCAAAGG 3008

Search completed: October 3, 2003, 05:18:13
Job time : 2736 secs

```



```

; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 111629
; LENGTH: 2652
; TYPE: DNA
; ORGANISM: Human
; US-10-027-632-111629

Query Match
Best Local Similarity 100.0%; Score 19; DB 13; Length 2652;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 1989 GCGTGGGCTTCTGGGCTC 2007
      |||
DB 1748 GCGTGGGCTTCTGGGCTC 1730

RESULT 43
US-10-027-632-111630/c
; Sequence 111630, Application US/10027632
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 111630
; LENGTH: 2652
; TYPE: DNA
; ORGANISM: Human
; US-10-027-632-111630

Query Match
Best Local Similarity 100.0%; Score 19; DB 13; Length 2652;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 1989 GCGTGGGCTTCTGGGCTC 2007
      |||
DB 1748 GCGTGGGCTTCTGGGCTC 1730

RESULT 44
US-10-128-714-386/c
; Sequence 386, Application US/0128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wengji
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroshtkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 386
; LENGTH: 3608
; TYPE: DNA
; ORGANISM: Aspergillus fumigatus
; US-10-128-714-386

Query Match
Best Local Similarity 100.0%; Score 19; DB 14; Length 3608;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 1570 GCGATGCCAGGATCAAGG 1586
      |||
DB 3026 GCGATGCCAGGATCAAGG 3008

RESULT 45
US-10-128-714-5386/c
; Sequence 5386, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wengji
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroshtkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362

```

```

; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 111629
; LENGTH: 2652
; TYPE: DNA
; ORGANISM: Human
; ORGANISM: Human
US-10-027-632-111629

Query Match
Best Local Similarity 0.7%; Score 19; DB 13; Length 2652;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 1989 GCGTGGGCTTCTGGGCTC 2007
Db 1748 GCGTGGGCTTCTGGGCTC 1730

RESULT 43
US-10-027-632-111630/c
; Sequence 111630, Application US/10027632
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 111630
; LENGTH: 2652
; TYPE: DNA
; ORGANISM: Human
; ORGANISM: Human
US-10-027-632-111630

Query Match
Best Local Similarity 0.7%; Score 19; DB 13; Length 2652;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 1989 GCGTGGGCTTCTGGGCTC 2007
Db 1748 GCGTGGGCTTCTGGGCTC 1730

RESULT 44
US-10-128-714-386/c
; Sequence 386, Application US/0128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wengji
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroszhkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 386
; LENGTH: 3608
; TYPE: DNA
; ORGANISM: Aspergillus fumigatus
; ORGANISM: Aspergillus fumigatus
US-10-128-714-386

Query Match
Best Local Similarity 0.7%; Score 19; DB 14; Length 3608;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 1570 GCGATGCCAGGATCAAGG 1586
Db 3026 GCGATGCCAGGATCAAGG 3008

RESULT 45
US-10-128-714-5386/c
; Sequence 5386, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wengji
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroszhkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362

```

```
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-3872

Query Match          0.7%; Score 19; DB 10; Length 1363;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1101 CCAACAAATTGAACAAG 1119
    |||||
Db 720 CCAACAAATTGAACAAG 702

RESULT 39
US-10-240-965-152/c
; Sequence 152, Application US/10240965
; Publication No. US20030165924A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: SHIFFMAN, Dov
; APPLICANT: SOMOGVI, Roland
; APPLICANT: LAWN, Richard M.
; APPLICANT: SEILHAMER, Jeffrey J.
; APPLICANT: PORTER, Gordon J.
; APPLICANT: MIKITA, Thomas
; APPLICANT: TAI, Julie
; TITLE OF INVENTION: GENES EXPRESSED IN FOAM CELL DIFFERENTIATION
; FILE REFERENCE: PA-0025 PCT
; CURRENT APPLICATION NUMBER: US/10/240,965
; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: 60/195,106
; PRIOR FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 276
; SOFTWARE: PERL Program
; SEQ ID NO 152
; LENGTH: 2341
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030165924A1 245099.8
US-10-240-965-152

Query Match          0.7%; Score 19; DB 12; Length 2341;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2025 AACGGGTTCGCCAGGAGAA 2043
    |||||
Db 104 AACGGGTTCGCCAGGAGAA 86

RESULT 40
US-10-125-237-52/c
; Sequence 52, Application US/10125237
; Publication No. US20030022329A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenchua
; APPLICANT: Zhou, Ping
; APPLICANT: Asundi, Vinod
; APPLICANT: Ren, Feiyan
; APPLICANT: Zhao, Qing A.
; APPLICANT: Xue, Aidong J.
; APPLICANT: Zhang, Jie
; APPLICANT: Wehrman, Tom
; APPLICANT: Wang, Jian-Rui
; TITLE OF INVENTION: No. US20030022329A1el Nucleic Acids and
; FILE REFERENCE: 791CIP2ADiv
; CURRENT APPLICATION NUMBER: US/10/125,237
; CURRENT FILING DATE: 2002-04-17
```

```
; PRIOR APPLICATION NUMBER: 09/668,317
; PRIOR FILING DATE: 2000-09-22
; PRIOR APPLICATION NUMBER: 09/552,929
; PRIOR FILING DATE: 2000-04-18
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: pt_FL_genes Version 2.0
; SEQ ID NO 52
; LENGTH: 2346
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (86) .. (1564)
US-10-125-237-52

Query Match          0.7%; Score 19; DB 14; Length 2346;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2025 AACGGGTTCGCCAGGAGAA 2043
    |||||
Db 111 AACGGGTTCGCCAGGAGAA 93

RESULT 41
US-10-105-891-52/c
; Sequence 52, Application US/10105891
; Publication No. US20030073099A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenchua
; APPLICANT: Zhou, Ping
; APPLICANT: Asundi, Vinod
; APPLICANT: Ren, Feiyan
; APPLICANT: Zhao, Qing A.
; APPLICANT: Xue, Aidong J.
; APPLICANT: Zhang, Jie
; APPLICANT: Wehrman, Tom
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Drmanac, Radoje
; TITLE OF INVENTION: No. US20030073099A1e: Nucleic Acids and
; FILE REFERENCE: 791CIP2A
; CURRENT APPLICATION NUMBER: US/10/105,891
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER: 09/668,317
; PRIOR FILING DATE: 2000-09-22
; PRIOR APPLICATION NUMBER: 05/552,929
; PRIOR FILING DATE: 2000-04-18
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: pt_FL_genes Version 2.0
; SEQ ID NO 52
; LENGTH: 2346
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (86) .. (1564)
US-10-105-891-52

Query Match          0.7%; Score 19; DB 14; Length 2346;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2025 AACGGGTTCGCCAGGAGAA 2043
    |||||
Db 111 AACGGGTTCGCCAGGAGAA 93

RESULT 42
US-10-027-632-111629/c
; Sequence 11629, Application US/10027632
; GENERAL INFORMATION:
```

```
QY 2297 CTGGGATCCAGGTCCTCA 2315
; ORGANISM: Homo Sapien
; US-10-184-644-269
;
Query Match 0.7%; Score 19; DB 14; Length 1300;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 595 CTGGGATCCAGGTCCTCA 613

RESULT 35
US-10-027-632-123666
; Sequence 269, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 123666
; LENGTH: 1188
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-123666

Query Match 0.7%; Score 19; DB 13; Length 1188;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2297 CTGGGATCCAGGTCCTCA 2315
; ORGANISM: Homo Sapien
; US-10-184-644-269/c
; Sequence 269, Application US/10184644
; Publication No. US2003004930A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P34301C227
; CURRENT APPLICATION NUMBER: US/10/184,644
; CURRENT FILING DATE: 2002-06-28
; PRIOR APPLICATION NUMBER: US 60/184,644
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 269
; LENGTH: 1300
; TYPE: PRT
US-10-184-644-269/c

Query Match 0.7%; Score 19; DB 14; Length 1300;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2025 AACGGGTTCCCGAGGAGAA 2043
; ORGANISM: Homo Sapien
; US-10-184-634-269/c
; Sequence 269, Application US/10184634
; Publication No. US2003068684A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P34301C217
; CURRENT APPLICATION NUMBER: US/10/184,634
; CURRENT FILING DATE: 2002-06-28
; PRIOR APPLICATION NUMBER: US 60/184,634
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 269
; LENGTH: 1300
; TYPE: PRT
US-10-184-634-269/c

Query Match 0.7%; Score 19; DB 14; Length 1300;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2025 AACGGGTTCCCGAGGAGAA 2043
; ORGANISM: Homo Sapien
; US-10-184-634-269
; Sequence 269, Application US/09938842A
; Publication No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zou, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIPT-300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 3872
; LENGTH: 1363
; TYPE: PRT
US-09-938-842A-3872/c
```

```
; APPLICANT: Dougan, Gordon
; APPLICANT: Wilson, Rebecca Ketry
; TITLE OF INVENTION: Genes and Proteins, and Their Use
; FILE REFERENCE: GJE-70C1
; CURRENT APPLICATION NUMBER: US/10/096,162
; CURRENT FILING DATE: 2002-11-10
; PRIOR APPLICATION NUMBER: GB 0105922.9
; PRIOR FILING DATE: 2001-03-09
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1146
; TYPE: DNA
; ORGANISM: group B streptococcus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1146)
; OTHER INFORMATION:
US-10-096-162-1
```

```
Query Match 0.7%; Score 19; DB 14; Length 1146;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 2606 CAGGAGAAATGGGAAAA 2624
|||||
DB 556 CAGGAGAAATGGGAAAA 574
```

```
RESULT 32
US-10-027-632-117674
; Sequence 117674, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 117674
; LENGTH: 1188
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-117674
```

```
Query Match 0.7%; Score 19; DB 13; Length 1188;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 2297 CTGGGATCCAGTCTCTCA 2315
|||||
DB 595 CTGGGATCCAGTCTCTCA 613
```

```
RESULT 33
US-10-027-632-117675
; Sequence 117675, Application US/10027632
```

```
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 117675
; LENGTH: 1188
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-117675
```

```
Query Match 0.7%; Score 19; DB 13; Length 1188;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 2297 CTGGGATCCAGTCTCTCA 2315
|||||
DB 595 CTGGGATCCAGTCTCTCA 613
```

```
RESULT 34
US-10-027-632-117676
; Sequence 117676, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 117676
; LENGTH: 1188
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-117676
```

```
Query Match 0.7%; Score 19; DB 13; Length 1188;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

; OTHER INFORMATION: n equals a,t,g, or c  
US-09-783-590-8616

Query Match 0.7%; Score 19; DB 10; Length 381;  
Best Local Similarity 100.0%; Pred. No. 73;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2310 TCCTCAGGACCCCGAGT 2328  
|||||  
Db 137 TCCTCAGGACCCCGAGT 119

## RESULT 28

US-09-918-995-30341/C  
; Sequence 30341, Application US/09918995  
; Publication No. US20030073623A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; FROM VARIOUS CDNA LIBRARIES  
; FILE REFERENCE: 20411-756  
; CURRENT APPLICATION NUMBER: US/09/918,995  
; CURRENT FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US/09/235,076  
; PRIOR FILING DATE: 1999-01-20  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 30341  
; LENGTH: 464  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: 11... (464)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-918-995-30341

Query Match 0.7%; Score 19; DB 11; Length 464;  
Best Local Similarity 100.0%; Pred. No. 73;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2025 AACGGTCCCCCAGGAA 2043  
|||||  
Db 113 AACGGTCCCCCAGGAA 95

## RESULT 29

US-10-027-632-28811/C  
; Sequence 28811, Application US/10027632  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 28811

; LENGTH: 580  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-28811

Query Match 0.7%; Score 19; DB 13; Length 580;  
Best Local Similarity 100.0%; Pred. No. 73;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 861 AAGATTAAAGTCAAGAA 879  
|||||  
Db 396 AAGATTAAAGTCAAGAA 378

## RESULT 30

US-09-814-353-20454  
; Sequence 20454, Application US/09814353  
; Publication No. US20030165831A1  
; GENERAL INFORMATION:  
; APPLICANT: Thompson, Pamela  
; APPLICANT: Lillie, James  
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR  
; IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
; THERAPY OF OVARIAN CANCER  
; FILE REFERENCE: MRI-006B  
; CURRENT APPLICATION NUMBER: US/09/814,353  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: US 60/191,031  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: US 60/207,124  
; PRIOR FILING DATE: 2000-05-25  
; PRIOR APPLICATION NUMBER: US 60/211,940  
; PRIOR FILING DATE: 2000-06-15  
; PRIOR APPLICATION NUMBER: US 60/216,820  
; PRIOR FILING DATE: 2000-07-07  
; PRIOR APPLICATION NUMBER: US 60/220,661  
; PRIOR FILING DATE: 2000-07-25  
; PRIOR APPLICATION NUMBER: US 60/257,672  
; PRIOR FILING DATE: 2000-12-21  
; NUMBER OF SEQ ID NOS: 22037  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 20454  
; LENGTH: 898  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: 1, 2, 3, 4, 5, 6, 7, 8  
; OTHER INFORMATION: n = A,T,C or G  
US-09-814-353-20454

Query Match 0.7%; Score 19; DB 12; Length 898;  
Best Local Similarity 100.0%; Pred. No. 74;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 864 ATTAAAGTCAAGAAAT 882  
|||||  
Db 716 ATTAAAGTCAAGAAAT 734

## RESULT 31

US-10-096-162-1  
; Sequence 1, Application US/10096162  
; Publication No. US2003010400CA1  
; GENERAL INFORMATION:  
; APPLICANT: Hughes, Martin John Glenton  
; APPLICANT: Santangelo, Joseph David  
; APPLICANT: Lane, Jonathan Douglas  
; APPLICANT: Feldman, Robert  
; APPLICANT: Moore, Jeanne Christine  
; APPLICANT: Dobson, Richard James  
; APPLICANT: Everest, Paul

LENGTH: 378361  
TYPE: DNA  
ORGANISM: Human  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1)...(378361)  
OTHER INFORMATION: n = A,T,C or G  
US-09-901-136-3

Query Match 0.7%; Score 20; DB 11; Length 378361;  
Best Local Similarity 100.0%; Pred.No. 26;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2123 AAAGGGAGAAATGGAAGA 2142  
DB 10250 AAAGGGAGAAATGGAAGA 10231

RESULT 26  
US-09-263-959-1  
Sequence 1, Application US/09263959  
Patent No. US20020150891A1  
GENERAL INFORMATION:  
APPLICANT: Hood, Leroy E.  
APPLICANT: Koop, Ben F.  
TITLE OF INVENTION: DIAGNOSTIC AND THERAPEUTIC COMPOSITIONS AND METHODS WHICH UTILIZE  
NUMBER OF SEQUENCES: 1279  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Seed and Berry LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: US  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/263,959  
FILING DATE: 05-MAR-1999  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Mcmasters, David E.  
REGISTRATION NUMBER: 33,963  
REFERENCE/DOCKET NUMBER: 920010.426C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 684973 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-263-959-1  
Query Match 0.7%; Score 20; DB 10; Length 684973;  
Best Local Similarity 100.0%; Pred.No. 27;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1842 CCAAAGGGAGAAATGGGC 1861  
DB 181777 CCAAAGGGAGAAATGGGC 181796  
RESULT 27  
US-09-783-590-8616/c  
Sequence 8616, Application US/09783590  
Patent No. US20020110850A1  
GENERAL INFORMATION:

APPLICANT: Dillon, Patrick J.  
APPLICANT: Haseltine, William A.  
ORGANISM: Li, Haodong  
APPLICANT: Rosen, Craig A.  
APPLICANT: Ruben, Steven M.  
TITLE OF INVENTION: Human Genes, Sequences, and Expression Products  
FILE REFERENCE: PO-16,201  
CURRENT APPLICATION NUMBER: US/09/783,590  
CURRENT FILING DATE: 2000-02-15  
PRIOR APPLICATION NUMBER: 08/420,856  
PRIOR FILING DATE: 1995-04-12  
PRIOR APPLICATION NUMBER: 08/346,731  
PRIOR FILING DATE: 1994-11-21  
NUMBER OF SEQ ID NOS: 12485  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 8616  
LENGTH: 381  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (27)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (37)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (38)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (58)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (72)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (155)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (178)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (210)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (222)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (257)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (259)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (313)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (330)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (340)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (355)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (361)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (365)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (368)

```
;
;
; NUMBER OF SEQUENCES: 982
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/070,927A
; FILING DATE: 04-May-2000
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/046,655
; FILING DATE: 1997-05-16
; APPLICATION NUMBER: 60/044,031
; FILING DATE: 1997-05-06
; APPLICATION NUMBER: 60/066,009
; FILING DATE: 1997-11-14
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenley K. Hoover
; REGISTRATION NUMBER: 40,302
; REFERENCE/DOCKET NUMBER: PB369
;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
;
; INFORMATION FOR SEQ ID NO: 621:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1635 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
; SEQUENCE DESCRIPTION: SEQ ID NO: 621:
US-09-070-927A-621

Query Match 0.7%; Score 20; DB 10; Length 1635;
Best Local Similarity 100.0%; Pred.No.23;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 38 TGCTGCTTGTTCAGAAATCT 57
| | | | | | | | | | | | | | | | | | | | | |
DB 766 TGCTGCTTGTTCAGAAATCT 785

RESULT 23
US-10-027-632-256103
; Sequence 256103, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/199,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09

;
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 256103
; LENGTH: 1816
; TYPE: DNA
; ORGANISM: Human
; US-10-027-632-256103

Query Match 0.7%; Score 20; DB 13; Length 1816;
Best Local Similarity 100.0%; Pred.No.23;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 717 AATTAATAAGGTTTAAAGAAA 736
| | | | | | | | | | | | | | | | | | | | | |
DB 1388 AATTAATAAGGTTTAAAGAAA 1407

RESULT 24
US-09-822-849A-13/c
; Sequence 13, Application US/09822849A
; Patent No. US20020045170A1
; GENERAL INFORMATION:
; APPLICANT: Wong, Gordon G.
; APPLICANT: Clark, Hilary
; APPLICANT: Agostino, Michael J.
; APPLICANT: Howes, Steven H.
; APPLICANT: Resnick, Richard J.
; APPLICANT: Gulukota, Kamalakari
; APPLICANT: Graham, James R.
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
; FILE REFERENCE: GIN 6403
; CURRENT APPLICATION NUMBER: US/09/822,849A
; CURRENT FILING DATE: 2001-09-04
; PRIOR APPLICATION NUMBER: 60/195,582
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 598
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 2493
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1204-
; OTHER INFORMATION: n = a,c,t, c, g
US-09-822-849A-13

Query Match 0.7%; Score 20; DB 9; Length 2493;
Best Local Similarity 100.0%; Pred.No.23;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2476 ATTCCTGGCCACCTGGTCC 2495
| | | | | | | | | | | | | | | | | | | | | |
DB 564 ATTCCTGGCCACCTGGTCC 545

RESULT 25
US-09-901-136-3/c
; Sequence 3, Application US/09901136
; Publication No. US20030039968A1
; GENERAL INFORMATION:
; APPLICANT: HU, Song et al.
; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
; FILE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL001273
; CURRENT APPLICATION NUMBER: US/09/901,136
; CURRENT FILING DATE: 2001-07-10
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
```

Publication No. US20030157568A1  
GENERAL INFORMATION:  
APPLICANT: Zuker, Charles S.  
APPLICANT: Adler, Jon Elliot  
APPLICANT: Ryba, Nick  
APPLICANT: Mueller, Ken  
APPLICANT: Hoon, Mark  
APPLICANT: The Regents of the University of California  
APPLICANT: The Government of the United States of America  
APPLICANT: as represented by the Secretary of the  
APPLICANT: Department of Health and Human Services  
TITLE OF INVENTION: SP, a No. US20030157568A1el Family of Taste Receptors  
FILE REFERENCE: 02307E-098000US  
CURRENT APPLICATION NUMBER: US/10/383,982  
CURRENT FILING DATE: 2003-03-07  
PRIOR APPLICATION NUMBER: US/09/393,634  
PRIOR FILING DATE: 1999-09-10  
NUMBER OF SEQ ID NOS: 92  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 61  
LENGTH: 954  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: human GR14  
US-10-383-982-61

Query Match 0.7%; Score 20; DB 12; Length 954;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 29 TGGTTTGGTGGCTGCTCTT 48  
Db 386 TGGTTTGGTGGCTGCTCTT 405

## RESULT 20

US-09-815-242-3957  
Sequence 3957, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haseibeck, Robert  
APPLICANT: Ohlsen, Kari L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 3957  
LENGTH: 1287  
TYPE: DNA

0  
ORGANISM: Enterococcus faecalis  
US-09-815-242-3957

Query Match 0.7%; Score 20; DB 9; Length 1287;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 38 TGCTGCTTCTTCAGATTCT 57  
Db 726 TGCTGCTTCTTCAGATTCT 745

## RESULT 21

US-09-815-242-6757  
Sequence 6757, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haseibeck, Robert  
APPLICANT: Ohlsen, Kari L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 6757  
LENGTH: 1299  
TYPE: DNA  
ORGANISM: Enterococcus faecalis  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)...(1299)  
US-09-815-242-6757

Query Match 0.7%; Score 20; DB 9; Length 1299;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 38 TGCTGCTTCTTCAGATTCT 57  
Db 726 TGCTGCTTCTTCAGATTCT 745

## RESULT 22

US-09-070-927A-621  
Sequence 621, Application US/09070927A  
Patent No. US2002020116A1  
GENERAL INFORMATION:  
APPLICANT: Charles A. Kunsch  
APPLICANT: Patrick J. Dillon  
APPLICANT: Steven Barash  
TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides



```
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 6505
; LENGTH: 577
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-6505

Query Match      0.7%; Score 20; DB 10; Length 577;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2505 AGAGGGTCCCGAGAGATTAC 2524
    |||||
Db 345 AGAGGGTCCCGAGAGATTAC 364

RESULT 16
US-09-263-959-372
; Sequence 372, Application US/09263959
; Patent No. US20020150891A1
; GENERAL INFORMATION:
; APPLICANT: Hood, Leroy E.
; APPLICANT: Koop, Ben F.
; TITLE OF INVENTION: DIAGNOSTIC AND THERAPEUTIC COMPOSITIONS AND METHODS WHICH UTI
; NUMBER OF SEQUENCES: 1279
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: US
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/263,959
; FILING DATE: 05-MAR-1999
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: McMasters, David D.
; REGISTRATION NUMBER: 33,963
; REFERENCE/DOCKET NUMBER: 920010.426C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 372:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 701 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-263-959-372

Query Match      0.7%; Score 20; DB 10; Length 701;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1842 CCAAAAGGAGAAATTGGC 1861
    |||||
Db 221 CCAAAAGGAGAAATTGGC 240

RESULT 17
US-09-393-982-61
; Sequence 61, Application US/09393634
```

```
; Patent No. US20020051997A1
; GENERAL INFORMATION:
; APPLICANT: Zuker, Charles S.
; APPLICANT: Adler, Jon Elliot
; APPLICANT: Ryba, Nick
; APPLICANT: Mueller, Ken
; APPLICANT: Hoon, Mark
; APPLICANT: The Regents of the University of California
; APPLICANT: The Government of the United States of America
; APPLICANT: as represented by the Secretary of the
; APPLICANT: Department of Health and Human Services
; TITLE OF INVENTION: SF, a No. US20020051997A1e1 Family of Taste Receptors
; FILE REFERENCE: 02307E-098000US
; CURRENT APPLICATION NUMBER: US/09/393,634
; CURRENT FILING DATE: 1999-09-10
; NUMBER OF SEQ ID NOS: 92
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 61
; LENGTH: 954
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: human GR14
US-09-393-634e61

Query Match      0.7%; Score 20; DB 9; Length 954;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 TCGTTTGGTGCTGCTTCTT 48
    |||||
Db 386 TCGTTTGGTGCTGCTTCTT 405

RESULT 18
US-09-510-332-27
; Sequence 27, Application US/09510332
; Publication No. US20030022278A1
; GENERAL INFORMATION:
; APPLICANT: Zuker, Charles S.
; APPLICANT: Adler, Jon Elliot
; APPLICANT: Ryba, Nick
; APPLICANT: Mueller, Ken
; APPLICANT: Hoon, Mark
; APPLICANT: The Regents of the University of California
; FILE REFERENCE: 02307E-098010US
; CURRENT APPLICATION NUMBER: US/09/510,332
; CURRENT FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 09/393,634
; PRIOR FILING DATE: 1999-09-10
; NUMBER OF SEQ ID NOS: 172
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 27
; LENGTH: 954
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: huamn T2R14 (hGR14)
US-09-510-332-27

Query Match      0.7%; Score 20; DB 11; Length 954;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 TCGTTTGGTGCTGCTTCTT 48
    |||||
Db 386 TCGTTTGGTGCTGCTTCTT 405

RESULT 19
US-10-383-982-61
; Sequence 61, Application US/10383982
```

US-09-854-133-174/c  
; Sequence 174, Application US/09854133  
; Publication No. US20020183499A1  
; GENERAL INFORMATION:  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Mohamath, Raodch  
; APPLICANT: Henderson, Robert A.  
; APPLICANT: Benson, Darin R.  
; APPLICANT: Secrist, Heather  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
; FILE REFERENCE: 210121.475C10  
; CURRENT APPLICATION NUMBER: US/09/854,133  
; CURRENT FILING DATE: 2001-05-11  
; NUMBER OF SEQ ID NOS: 735  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 174  
; LENGTH: 548  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-854-133-174

Query Match 0.7%; Score 20; DB 9; Length 479;  
Best Local Similarity 100.0%; Pred. No. 22;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 29 TGGTTTGGTGGCTCTCTT 48  
|||||  
DB 386 TGGTTTGGTGGCTCTCTT 367  
|||||

US-09-864-761-10823  
; OTHER INFORMATION: MAP TO AC006518.17  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.94  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1  
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.99  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.96  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.2

Query Match 0.7%; Score 20; DB 9; Length 479;  
Best Local Similarity 100.0%; Pred. No. 22;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 TGGTTTGGTGGCTCTCTT 48  
|||||  
DB 386 TGGTTTGGTGGCTCTCTT 367  
|||||

RESULT 12  
US-09-738-973-174/c  
; Sequence 174, Application US/09738973  
; Patent No. US2002010563A1  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Henderson, Robert A.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Fling, Steven P.  
; APPLICANT: Mohamath, Raodch  
; APPLICANT: Algate, Paul A.  
; APPLICANT: Secrist, Heather  
; APPLICANT: Indrias, Carol Yoseph  
; APPLICANT: Benson, Darin R.  
; APPLICANT: Elliot, Mark  
; APPLICANT: Mannion, Jane  
; APPLICANT: Kalos, Michael D.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
; FILE REFERENCE: 210121.475C9  
; CURRENT APPLICATION NUMBER: US/09/738,973  
; CURRENT FILING DATE: 2000-12-14  
; NUMBER OF SEQ ID NOS: 587  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 174  
; LENGTH: 548  
; TYPE: DNA  
; ORGANISM: Homo sapien  
US-09-738-973-174

Query Match 0.7%; Score 20; DB 10; Length 548;  
Best Local Similarity 100.0%; Pred. No. 22;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2476 ATTCCTGGCCACCTGGTCC 2495  
|||||  
DB 543 ATTCCTGGCCACCTGGTCC 524  
|||||

RESULT 13

US-09-854-133-174/c  
; Sequence 174, Application US/09854133  
; Publication No. US20020183499A1  
; GENERAL INFORMATION:  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Mohamath, Raodch  
; APPLICANT: Henderson, Robert A.  
; APPLICANT: Benson, Darin R.  
; APPLICANT: Secrist, Heather  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
; FILE REFERENCE: 210121.475C10  
; CURRENT APPLICATION NUMBER: US/09/854,133  
; CURRENT FILING DATE: 2001-05-11  
; NUMBER OF SEQ ID NOS: 735  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 174  
; LENGTH: 548  
; TYPE: DNA  
; ORGANISM: Homo sapien  
US-09-854-133-174

Query Match 0.7%; Score 20; DB 10; Length 548;  
Best Local Similarity 100.0%; Pred. No. 22;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2476 ATTCCTGGCCACCTGGTCC 2495  
|||||  
DB 543 ATTCCTGGCCACCTGGTCC 524  
|||||

RESULT 14  
US-10-144-649A-174/c  
; Sequence 174, Application US/10144649A  
; Publication No. US20030118599A1  
; GENERAL INFORMATION:  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Wang, Tongtong  
; APPLICANT: Pan, Liqun  
; APPLICANT: Algate, Paul A.  
; APPLICANT: McNeill, Patricia D.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
; FILE REFERENCE: 210121.475C11  
; CURRENT APPLICATION NUMBER: US/10/144,649A  
; CURRENT FILING DATE: 2002-08-21  
; NUMBER OF SEQ ID NOS: 749  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 174  
; LENGTH: 548  
; TYPE: DNA  
; ORGANISM: Homo sapien  
US-10-144-649A-174

Query Match 0.7%; Score 20; DB 14; Length 548;  
Best Local Similarity 100.0%; Pred. No. 22;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2476 ATTCCTGGCCACCTGGTCC 2495  
|||||  
DB 543 ATTCCTGGCCACCTGGTCC 524  
|||||

RESULT 15  
US-09-867-70k-6505  
; Sequence 6505, Application US/09867701  
; Patent No. US20020132237A1  
; GENERAL INFORMATION:  
; APPLICANT: Aglate, Paul A.  
; APPLICANT: Jones, Robert  
; APPLICANT: Harlocker, Susan J.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
; FILE REFERENCE: 210121.475C11  
; CURRENT APPLICATION NUMBER: US/09/867,701  
; CURRENT FILING DATE: 2000-12-14  
; NUMBER OF SEQ ID NOS: 587  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 174  
; LENGTH: 548  
; TYPE: DNA  
; ORGANISM: Homo sapien  
US-09-867-70k-6505

```
; Sequence 16328, Application US/09864761;
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 16328
; LENGTH: 466
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL034452.8
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.63
US-09-864-761-16328

Query Match 1.0%; Score 28; DB 9; Length 466;
Best Local Similarity 100.0%; Pred. No. 0.0017;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 2573 GACGTCCAGGTGTCCAGGATTAAGG 2600
|||||
Db 466 GACGTCCAGGTGTCCAGGATTAAGG 439

RESULT 10
US-09-938-842A-2425
; Sequence 2425, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Harper, Jeff
; APPLICANT: Krieps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAMR, AND METHODS OF USE
; FILE REFERENCE: SCRIIP300-3
; CURRENT APPLICATION NUMBER: US/39/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,966
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 2425
; LENGTH: 393
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-2425

Query Match 0.7%; Score 20; DB 10; Length 393;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 746 TTTCACCAAAAAGATATAA 765
|||||
Db 33 TTTCACCAAAAAGATATAA 52

RESULT 11
US-09-864-761-10823/c
; Sequence 10823, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
```

```
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
; SEQ ID NO 31680
; LENGTH: 181
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL034452.8
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 29
; OTHER INFORMATION: EST HUMAN HIT: N46880.1, EVALUE 1.00e-38
; OTHER INFORMATION: NT HIT: AF121217.1, EVALUE 1.70e+00
US-09-864-761-31680
```

```
Query Match 2.9%; Score 82; DB 9; Length 181;
Best Local Similarity 100.0%; Pred. No. 2.2e-31;
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2598 AGGCCTACCGAGGAAGTGGGAAAAGGAGCCAGAGGTTTGGGTATCTCGAGACA 2657
Db 165 AGGCCTACCGAGGAAGTGGGAAAAGGAGCCAGAGGTTTGGGTATCTCGAGACA 106

QY 2658 AGGTCCTCTGGTCCCCCAGGT 2679
Db 105 AGGTCCTCTGGTCCCCCAGGT 84
```

```
RESULT 7
US-10-027-632-124647/c
; Sequence 124647, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027.632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
```

```
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 924647
; LENGTH: 1197
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-124647
```

```
Query Match 2.0%; Score 57; DB 13; Length 1197;
Best Local Similarity 100.0%; Pred. No. 1.8e-18;
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1955 AGGAAGCAAGGTGAACCTGGAATTCAGGGATGCTGGGGCTTGTGGCTCAAGG 2811
Db 65 AGGAAGCAAGGTGAACCTGGAATTCAGGGATGCTGGGGCTTGTGGCTCAAGG 9
```

```
RESULT 8
US-10-027-632-197684/c
; Sequence 197684, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027.632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 197684
; LENGTH: 663
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-197684
```

```
Query Match 1.6%; Score 46; DB 13; Length 663;
Best Local Similarity 100.0%; Pred. No. 8.6e-13;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1894 GGAATCCAGGATTCCTCGAAACCGAGGATTAATGGCCCAAGG 1849
Db 233 GGAATCCAGGATTCCTCGAAACCGAGGATTAATGGCCCAAGG 188

RESULT 9
US-09-864-761-16328/c
```

QY 2839 CCGTTCAGAAAGACCAACTATTAG 2865  
Db 325 CCGTTCAGAAAGACCAACTATTAG 351

## RESULT 4

US-09-918-995-33525  
; Sequence 33525, Application US/09918995  
; Publication No. US20030073623A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; FROM VARIOUS CDNA LIBRARIES  
; FILE REFERENCE: 20411-756  
; CURRENT APPLICATION NUMBER: US/09/918,995  
; CURRENT FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US/09/235,076  
; PRIOR FILING DATE: 1999-01-20  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 33525  
; LENGTH: 412  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)..(412)  
; OTHER INFORMATION: n = A,T,C or G

US-09-918-995-33525

Query Match 7.8%; Score 224; DB 11; Length 412;  
Best Local Similarity 100.0%; Pred. No. 1.2e-104;  
Matches 224; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1279 TGCCTTAATGTCCTCCAGTGATGATGTTCAACTCCAGTCCCTGATTTGTCCTCCGGA 1338  
Db 188 TGCCTTAATGTCCTCCAGTGATGATGTTCAACTCCAGTCCCTGATTTGTCCTCCGGA 247  
QY 1339 AAACGAGACTTCAAGGCCCAAGGTCACCTGGACTGCTGGGACCTGGGTACCT 1393  
Db 248 AAACGAGACTTCAAGGCCCAAGGTCACCTGGACTGCTGGGACCTGGGTACCT 307  
QY 1399 GGACAACTGTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1458  
Db 308 GGACAACTGTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 367  
QY 1459 CCAGGATCTCCAGGAATACAGGAGCTCGAGGACTACAGGTTA 1502  
Db 368 CCAGGATCTCCAGGAATACAGGAGCTCGAGGACTACAGGTTA 411

## RESULT 5

US-09-864-761-32810/c  
; Sequence 32810, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aomic-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04

; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,697  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49217  
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 32810  
; LENGTH: 221  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AL034452.8  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.63  
; OTHER INFORMATION: NT HIT: L02917.1, EVALUE 2.10e+00  
; OTHER INFORMATION: EST HUMAN HIT: AA451616.1, EVALUE 1.00e-103  
; OTHER INFORMATION: SWISSPROT HIT: C53653, EVALUE 2.10e+00  
US-09-864-761-32810

Query Match 5.8%; Score 166; DB 9; Length 221;  
Best Local Similarity 100.0%; Pred. No. 1e-74;  
Matches 166; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2435 GAATTTGTGATCATTTGCTGTGCTCCCAACATGCTCCCGGGTATTCTCGGCACCTGGTC 2494  
Db 221 GAATTTGTGATCATTTGCTGTGCTCCCAACATGCTCCCGGGTATTCTCGGCACCTGGTC 162  
QY 2495 CGATAGGCCACAGAGGTCCTCCAGAGATTACCTGGTTTCCAGGAAGAGATGGTGTTCCTG 2554  
Db 161 CGATAGGCCACAGAGGTCCTCCAGAGATTACCTGGTTTCCAGGAAGAGATGGTGTTCCTG 102  
QY 2555 GATTAGTGGTGTTCCTCGAGTCCAGTGTCCAGAGTTAAAGG 2600  
Db 101 GATTAGTGGTGTTCCTCGAGTCCAGTGTCCAGAGTTAAAGG 56

## RESULT 6

US-09-864-761-31680/c  
; Sequence 31680, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aomic-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23



90 18 0.6 557 13 US-10-027-632-85796 Sequence 85796, A  
91 18 0.6 557 13 US-10-027-632-320345 Sequence 320345,  
92 18 0.6 557 13 US-10-027-632-320346 Sequence 320346,  
93 18 0.6 557 13 US-10-027-632-320347 Sequence 320347,  
94 18 0.6 557 13 US-10-027-632-320348 Sequence 320348,  
95 18 0.6 570 13 US-10-027-632-115045 Sequence 115045,  
96 18 0.6 570 13 US-10-027-632-236526 Sequence 236526,  
97 18 0.6 581 13 US-10-027-632-70650 Sequence 70650, A  
98 18 0.6 583 13 US-10-027-632-73394 Sequence 73394, A  
99 18 0.6 583 13 US-10-027-632-189884 Sequence 189884,  
100 18 0.6 583 13 US-10-027-632-189885 Sequence 189885,

## ALIGNMENTS

RESULT 1  
US-09-764-864-232  
; Sequence 232, Application US/09764864  
; Patent No. US20020132753A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: P7223  
; CURRENT APPLICATION NUMBER: US/09/764,864  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 1792  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 232  
; LENGTH: 2375  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-764-864-232

Query Match 44.7%; Score 1282; DB 10; Length 2375;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1282; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1584 AAAGGTGAAATGGTGCCAAAGAGACAAAGATCACCTGGATTATGGCAAAAGGG 1643  
DB 9 AAAGGTGAAATGGTGCCAAAGAGACAAAGATCACCTGGATTATGGCAAAAGGG 58  
QY 1644 TGCAAAAGGTGAAAGGGGAATGCTGGCTTCCCTGGCCCTCCCTGGACCTGCTGGGAACC 1703  
DB 69 TGCAAAAGGTGAAAGGGGAATGCTGGCTTCCCTGGCCCTCCCTGGACCTGCTGGGAACC 128  
QY 1704 AGGAACATGGAAGATGGAATTAATGGTATGGTATGCTCCGGTTCAAGGGAGACGAGATC 1763  
DB 129 AGGAACATGGAAGATGGAATTAATGGTATGGTATGCTCCGGTTTCAAGGGAGACGAGATC 188  
QY 1764 CCCTGGTCTCCGGGGCAGGATGGAACACGGGGAGAGCCGGAAATCCAGGATTCCTGG 1823  
DB 189 CCCTGGTCTCCGGGGCAGGATGGAACACGGGGAGAGCCGGAAATCCAGGATTCCTGG 248  
QY 1824 AAACCCAGGATTAATGGGCCAAAGGAGAAATTTGGGCTCCAGGACAGCAAGAAAAA 1883  
DB 249 AAACCCAGGATTAATGGGCCAAAGGAGAAATTTGGGCTCCAGGACAGCAAGAAAAA 308  
QY 1884 AGAGCCCCCAGGATGCTGTTTAAATGGGAAGCAATGGCTCACCGCCAGCCTCGAAC 1943  
DB 309 AGGAGCCCCAGGATGCTGTTTAAATGGGAAGCAATGGCTCACCGCCAGCCTCGAAC 368  
QY 1944 ACCGGGATTAAGGGAAGCAAGGTAACCTTGAATTCAGGGGATGCTGGGGCTTCTGG 2203  
DB 369 ACCGGGATTAAGGGAAGCAAGGTAACCTTGAATTCAGGGGATGCTGGGGCTTCTGG 428  
QY 2004 GCTCAAGGGAGAACCCAGGACACCGGTTTCCCGAGGAAACCGAGTACATGGGTTTACC 2063  
DB 429 GCTCAAGGGAGAACCCAGGACACCGGTTTCCCGAGGAAACCGAGTACATGGGTTTACC 488  
QY 2064 CGGGATTCAAGGAAAAAGGGGACAAAGGAATCAAGGTGAAAGGTTATTCAGGGTCA 2123

DB 489 CCGGATTCAAGGAAAAAGGGGACAAAGGAATCAAGGTCAAAAAGGTATTCAAGGCTCA 548  
QY 2-24 AAAGGGAGAAAATCGAAGACAGAGGAAATTCAGGGCCACAGGGAATTCAGGCCATCATGG 2183  
DB 549 AAAGGGAGAAAATCGAAGACAGAGGAAATTCAGGGCCACAGGGAATTCAGGCCATCATGG 608  
QY 2-84<sup>6</sup> TGCAAAAGGAGAGAGAGGTGAAAGGGAGAACCTGGTGTCCGAGGTGCCATTGGATCAAA 2243  
DB 609 TGCAAAAGGAGAGAGAGGTGAAAGGGAGAACCTGGTGTCCGAGGTGCCATTGGATCAAA 668  
QY 2244 AGGAATATCTGGGGTGGATGCTTTGATGGGGCCCGCAGGTCTTAAGGGGCAACCTGGGGA 2303  
DB 669 AGGAATATCTGGGGTGGATGCTTTGATGGGGCCCGCAGGTCTTAAGGGGCAACCTGGGGA 728  
QY 2304 TCCAGGTCTCTCAGGACCCGCCAGGTTTGGATGGGAAGCCCGGAAGAGAGTTCAGAAC 2363  
DB 729 TCCAGGTCTCTCAGGACCCGCCAGGTTTGGATGGGAAGCCCGGAAGAGAGTTCAGAAC 788  
QY 2364 ATTTATTCGACAAAGTTTGCACAGATGTAATAAGAGCCACAGCTACCACTTTACTTCAGAG 2423  
DB 789 ATTTATTCGACAAAGTTTGCACAGATGTAATAAGAGCCACAGCTACCACTTTACTTCAGAG 848  
QY 2424 TGGAGAATTAGAATTTGTGATCATTTGCCTGTCCACATGGCTCCCGGGTATTCCTGG 2483  
DB 849 TGGAGAATTAGAATTTGTGATCATTTGCCTGTCCACATGGCTCCCGGGTATTCCTGG 908  
QY 2484 GCCACCTGGTCCGATAGCCCGCAGAGGGTCCAGAGGATTACCTGGTTTGCAGGAAGAGA 2543  
DB 909 GCCACCTGGTCCGATAGCCCGCAGAGGGTCCAGAGGATTACCTGGTTTGCAGGAAGAGA 968  
QY 2544 TGGTGTCTCGGATTAAGTGGGTGTCCTCGACGTCCAGGTGTGAGAGGATTAAGAGCCT 2603  
DB 963 TGGTGTCTCGGATTAAGTGGGTGTCCTCGACGTCCAGGTGTGAGAGGATTAAGAGCCT 1028  
QY 2604 ACCAGGAGAAATCGGGAAAGGGAGCCAGGGTTTGGGTATCTCCGAGAACAGGTCC 2663  
DB 1029 ACCAGGAGAAATCGGGAAAGGGAGCCAGGGTTTGGGTATCTCCGAGAACAGGTCC 1058  
QY 2664 TCCTGGTCTCCCGAGGTCCAGAGGCCCTCTCGAATAAGCAAAAGAGGTCTCTCCAGGAGA 2723  
DB 1089 TCCTGGTCTCCCGAGGTCCAGAGGCCCTCTCGAATAAGCAAAAGAGGTCTCTCCAGGAGA 1148  
QY 2724 CCGAGGTCTCCTCGCAAGATGGAGACCATGGAAAACCTGGGAATCCAAAGGCAACAGG 2783  
DB 1149 CCGAGGTCTCCTCGCAAGATGGAGACCATGGAAAACCTGGGAATCCAAAGGCAACAGG 1208  
QY 2784 CCCCCCAGGATCTCGACCCCATCACTATGTTTGTAGTGAATTCGCCAGAGAGATCCGTT 2843  
DB 1209 CCCCCCAGGATCTCGACCCCATCACTATGTTTGTAGTGAATTCGCCAGAGAGATCCGTT 1268  
QY 2844 CAGAAAGGACCAAACTATTAG 2865  
DB 1269 CAGAAAGGACCAAACTATTAG 1290

RESULT 2  
US-10-037-270-282  
; Sequence 282, Application US/10037270  
; Publication No. US20030104529A1  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Liu, Chenghua  
; APPLICANT: Asundi, Vinod  
; APPLICANT: Zhang, Jie  
; APPLICANT: Ren, Feiyan  
; APPLICANT: Chen, Rui-hong  
; APPLICANT: Zhao, Qing A.  
; APPLICANT: Wehrman, Tom  
; APPLICANT: Xue, Aidong J.  
; APPLICANT: Yang, Yonghong  
; APPLICANT: Wang, Jian-Rui  
; APPLICANT: Zhou, Ping  
; APPLICANT: Ma, Yuning

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2003, 23:50:00 ; Search time 2728 seconds  
(without alignments)  
2678.968 Million cell updates/sec

Title: US-09-996-611B-5

Perfect score: 2865

Sequence: 1 agggctcactattacatt.....gaaaggacccaactattag 2865

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 1708419 seqs, 1275411651 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3416839

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database :

Published Applications NA:\*

- 1: /cgn2\_6/prodata/1/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/prodata/1/pubpna/FCR\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/prodata/1/pubpna/US06\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/prodata/1/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/prodata/1/pubpna/US07\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/prodata/1/pubpna/FCRUS\_PUBCOMB.seq.\*
- 7: /cgn2\_6/prodata/1/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/prodata/1/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/prodata/1/pubpna/US09A\_PUBCOMB.seq.\*
- 10: /cgn2\_6/prodata/1/pubpna/US09B\_PUBCOMB.seq.\*
- 11: /cgn2\_6/prodata/1/pubpna/US09C\_PUBCOMB.seq.\*
- 12: /cgn2\_6/prodata/1/pubpna/US09\_NEW\_PUB.seq.\*
- 13: /cgn2\_6/prodata/1/pubpna/US10A\_PUBCOMB.seq.\*
- 14: /cgn2\_6/prodata/1/pubpna/US10B\_PUBCOMB.seq.\*
- 15: /cgn2\_6/prodata/1/pubpna/US10C\_NEW\_PUB.seq.\*
- 16: /cgn2\_6/prodata/1/pubpna/US60\_NEW\_PUB.seq.\*
- 17: /cgn2\_6/prodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMAR-ES

Result No.	Score	Query Match	Length	ID	Description
1	1282	44.7	2375	10	US-09-764-864-232
2	1062	37.1	2230	14	US-10-037-270-282
3	267	9.3	496	11	US-09-918-995-17490
4	224	7.8	412	11	US-09-918-995-33525
5	166	5.8	221	9	US-09-864-761-32810
6	82	2.9	181	9	US-09-864-761-31880
7	57	2.0	1197	13	US-10-027-632-124647
8	46	1.6	663	13	US-10-027-632-197684
9	28	1.0	466	9	US-09-864-761-16328
10	20	0.7	393	10	US-09-938-842A-2425
11	20	0.7	479	9	US-09-864-761-10823
12	20	0.7	548	10	US-09-738-973-174
13	20	0.7	548	10	US-10-144-649A-174
14	20	0.7	548	10	US-10-144-649A-174
15	20	0.7	577	10	US-09-867-701-6505
16	20	0.7	701	10	US-09-263-959-372

17	20	0.7	954	9	US-09-393-634-61	Sequence 61, Appl
18	20	0.7	954	11	US-09-510-332-27	Sequence 27, Appl
19	20	0.7	954	12	US-10-383-982-61	Sequence 61, Appl
20	23	0.7	1287	9	US-09-815-242-3957	Sequence 3957, Ap
21	20	0.7	1299	9	US-09-815-242-6757	Sequence 6757, Ap
22	20	0.7	1635	10	US-09-070-927A-621	Sequence 622, App
23	20	0.7	1816	13	US-10-027-632-256103	Sequence 256103,
c 24	20	0.7	2493	9	US-09-922-849A-13	Sequence 13, Appl
c 25	20	0.7	378361	11	US-09-901-136-3	Sequence 3, Appl
c 26	20	0.7	684973	10	US-09-263-959-1	Sequence 1, Appl
c 27	19	0.7	391	10	US-09-783-580-8616	Sequence 8616, Ap
c 28	19	0.7	464	11	US-09-928-995-30341	Sequence 30341, A
c 29	19	0.7	580	13	US-10-027-632-28811	Sequence 28811, A
c 30	19	0.7	998	12	US-09-814-353-20454	Sequence 20454, A
c 31	19	0.7	1146	14	US-10-096-162-1	Sequence 1, Appl
c 32	19	0.7	1188	13	US-10-027-632-117674	Sequence 117674,
c 33	19	0.7	1188	13	US-10-027-632-117675	Sequence 117675,
c 34	19	0.7	1188	13	US-10-027-632-117676	Sequence 117676,
c 35	19	0.7	1188	13	US-10-027-632-123666	Sequence 123666,
c 36	19	0.7	1300	14	US-10-184-644-269	Sequence 269, App
c 37	19	0.7	1300	14	US-10-184-644-269	Sequence 269, App
c 38	19	0.7	1363	10	US-09-938-842A-3872	Sequence 3872, Ap
c 39	19	0.7	2341	12	US-10-240-965-152	Sequence 152, App
c 40	19	0.7	2346	14	US-10-125-237-52	Sequence 52, Appl
c 41	19	0.7	2346	14	US-10-105-891-52	Sequence 52, Appl
c 42	19	0.7	2652	13	US-10-027-632-11629	Sequence 11629,
c 43	19	0.7	2652	13	US-10-027-632-11630	Sequence 11630,
c 44	19	0.7	3608	14	US-10-128-714-386	Sequence 386, App
c 45	19	0.7	3609	14	US-10-128-714-5386	Sequence 5386, Ap
c 46	19	0.7	66686	10	US-09-736-960-86	Sequence 86, Appl
c 47	18	0.6	25	14	US-10-098-283B-48341	Sequence 48341, A
c 48	18	0.6	115	9	US-09-815-242-3122	Sequence 3122, Ap
c 49	18	0.6	118	9	US-09-864-761-22670	Sequence 22670, A
c 50	18	0.6	139	9	US-09-815-242-3611	Sequence 3611, Ap
c 51	18	0.6	186	10	US-09-783-590-12193	Sequence 12193, A
c 52	18	0.6	261	9	US-09-815-242-3045	Sequence 3045, Ap
c 53	18	0.6	284	13	US-10-040-739-740	Sequence 740, App
c 54	18	0.6	304	13	US-10-040-739-468	Sequence 468, App
c 55	18	0.6	340	10	US-09-954-531-1136	Sequence 1136, Ap
c 56	18	0.6	340	10	US-09-954-531-1136	Sequence 1136, Ap
c 57	18	0.6	372	9	US-09-815-242-3153	Sequence 3153, Ap
c 58	18	0.6	377	10	US-09-736-457-1127	Sequence 1127, Ap
c 59	18	0.6	377	10	US-09-736-457-1127	Sequence 1127, Ap
c 60	18	0.6	377	10	US-09-902-941-1127	Sequence 1127, Ap
c 61	18	0.6	377	10	US-09-902-941-1127	Sequence 1127, Ap
c 62	18	0.6	377	10	US-09-849-626-1127	Sequence 1127, Ap
c 63	18	0.6	377	10	US-09-849-626-1127	Sequence 1127, Ap
c 64	18	0.6	377	12	US-10-113-872-1127	Sequence 1127, Ap
c 65	18	0.6	377	12	US-10-113-872-1127	Sequence 1127, Ap
c 66	18	0.6	377	14	US-10-017-754-1127	Sequence 1127, Ap
c 67	18	0.6	377	14	US-10-017-754-1127	Sequence 1127, Ap
c 68	18	0.6	394	13	US-10-027-632-52921	Sequence 52921, A
c 69	18	0.6	395	13	US-10-027-632-82160	Sequence 82160, A
c 70	18	0.6	395	13	US-10-027-632-302139	Sequence 302139,
c 71	18	0.6	429	11	US-09-918-995-5943	Sequence 5943, Ap
c 72	18	0.6	439	11	US-09-918-995-13940	Sequence 13940, A
c 73	18	0.6	441	10	US-09-954-456-337	Sequence 327, App
c 74	18	0.6	444	11	US-09-764-891-132	Sequence 132, App
c 75	18	0.6	457	11	US-09-918-995-21550	Sequence 21550, A
c 76	18	0.6	461	9	US-09-864-761-5910	Sequence 5910, Ap
c 77	18	0.6	467	9	US-09-770-444-259	Sequence 359, App
c 78	18	0.6	469	9	US-09-815-343-359	Sequence 359, App
c 79	18	0.6	474	9	US-09-864-761-113955	Sequence 13955, A
c 80	18	0.6	479	11	US-09-918-995-23654	Sequence 23654, A
c 81	18	0.6	515	9	US-09-864-761-7318	Sequence 7318, Ap
c 82	18	0.6	517	13	US-10-027-632-87326	Sequence 87326, A
c 83	18	0.6	517	13	US-10-027-632-367167	Sequence 367167,
c 84	18	0.6	531	11	US-09-862-540-18	Sequence 18, Appl
c 85	18	0.6	550	13	US-10-027-632-193269	Sequence 193269,
c 86	18	0.6	550	13	US-10-027-632-193270	Sequence 193270,
c 87	18	0.6	557	13	US-10-027-632-85793	Sequence 85793, A
c 88	18	0.6	557	13	US-10-027-632-85794	Sequence 85794, A
c 89	18	0.6	557	13	US-10-027-632-85795	Sequence 85795, A



Db	4346	CGGTCCTCAGGGACCTGCAGGAAAGCCTGGTCCAGA-----AGGTCCTTCGGGGC	4394
Qy	2447	ATTCCCTGTCCCAACATGGCTCCCGGGTATTCTCTGGGCCACCTGGTCCGATAGGCCCCAG	2506
Db	4395	ATCCCTGGTCC--TGTGGAGAACAGGTCTCCCTGGAGCTGCAGGCCAAGATGGACCAC	4452
Qy	2507	AGGTCCTCCAGAGGATTACCTGGTTTGCAGGAAGAGATGTTCTCTGGAATTAGTGGGTG	2566
Db	4453	CTGGTCTATGGGACCTCCTGGTTACCTGGTCTCAAGGTGACCTGGCTCCCAAGGGTG	4512
Qy	2567	TCCCTGGACGTCCAGGTCTCAGAGGATTAAAGGCCTACCAGGAGNAATGGGGAAAAAG	2626
Db	4513	AAAAGGACATCTCTGGTTTAAATTGGCTTGAATTGGTCTCTCCAGGAGAACAAAGGGGAAAAAG	4572
Qy	2627	GGAGCCCAAGGGTTTGGGTATCTCTGGAGAACAAAGGTCTCTCTGGTCCCTCCAGGTCCAGAGG	2686
Db	4573	GTGACCGAGGGCTC-----CCTGGAACTCAAGGATCTCCAGGAGCAAAAGGGGATGGGG	4626
Qy	2687	GCCCTCTCTGGAAATAGCAAGAGGTCTCTCAGAGACCCAGGTCTCCCTGGGCAAGATG	2746
Db	4627	GAATTCTCTGGTCTT-----GCTGGTCCCTTAGGTCCACCTGGTCTCTCCAGGCTTACCAG	4680
Qy	2747	GAGACCATGGAAAACTTGGAAATCCNAGGGCAACCAAGGCCCCCGCAGGC	2793
Db	4681	GTCTCAAGGCCCAAGGGTAAACAAAGGCTCTACTGGACCCGCTGGC	4727

Search completed: October 3, 2003, 00:00:11  
Job time : 715 secs

Db 4346 CGTCTCTCAGGACCTGCAGGAAAGCTGTGTCAGA-----AGTCTTTCGGGGC 4394  
Qy 2447 ATTGCTGTCTCAACATAGCTCCCGGGGTATTCTCTGGGCCACCTGTGTCCGATAGGCCAG 2506  
Db 4395 ATCCCTGGTCC--TGTGGGAAACAGAGTCTCTCTGGAGCTGCAGGCCAAGATGGACAC 4452  
Qy 2507 AGGTCCAGAGGATTACTCGTTTGCAGAAAGATGTTCTCTGGATTAGTGGTG 2566  
Db 4453 CTGTCTCTATGGGACCTCTCGCTTACCTGTCTCAAAGGTGACCTGGCTCCAAAGGTT 4512  
Qy 2567 TCCCTGGAGCTCCAGGCTCAGAGATTAAAGGCTTACCGAGAGAAATGGGGAAAAG 2626  
Db 4513 AAAAGGACATCTCGTTTAAATTGGCTGATTGTCTCTCAGGAGAACAGGGGAAAAG 4572  
Qy 2627 GGAGCCAAAGGTTTGGGTATCTCTGGAGAAAGGTCTCTCTGGTCCCGCCAGGTCCAGAG 2686  
Db 4573 GTGACCGAGGGCTC-----CCTGGACTCAAGATCTCCAGGAGCAAAAGGGATGGG 4626  
Qy 2687 GCCCTCTGGAATAGCAAGAGGTCTCAGAGACCCAGGTCTCTCTGGCAAGATG 2746  
Db 4627 GAATTCCTGTGTCCT-----GCTGTCTCTTAGGTCCACCTGTCTCTCCAGGCTTACAG 4680  
Qy 2747 GAGACCATGAAAACCTCGAATCAAGGGCAACCAAGGCCCGCCCGCAGGC 2793  
Db 4681 C:CTCAAGGCCCAAGGTTAAAGGCTCTACTGGAACCGCTGGC 4727

RESULT 15  
US-09-954-456-762  
; Sequence 762, Application US/09954456  
; Patent No. US20020115057A1  
; GENERAL INFORMATION:  
; APPLICANT: Young, Paul  
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Cand  
; FILE REFERENCE: 689290-76  
; CURRENT APPLICATION NUMBER: US/09/954,456  
; PRIOR FILING DATE: 2001-09-18  
; PRIOR APPLICATION NUMBER: US/60/233,617  
; PRIOR FILING DATE: 2000-09-18  
; PRIOR APPLICATION NUMBER: US/60/234,052  
; PRIOR FILING DATE: 2000-09-20  
; PRIOR APPLICATION NUMBER: US/60/234,923  
; PRIOR FILING DATE: 2000-09-25  
; PRIOR APPLICATION NUMBER: US/60/235,134  
; PRIOR FILING DATE: 2000-09-25  
; PRIOR APPLICATION NUMBER: US/60/235,637  
; PRIOR FILING DATE: 2000-09-26  
; PRIOR APPLICATION NUMBER: US/60/235,638  
; PRIOR FILING DATE: 2000-09-26  
; PRIOR APPLICATION NUMBER: US/60/235,711  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: US/60/235,720  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: US/60/235,840  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: US/60/235,863  
; PRIOR FILING DATE: 2000-09-27  
; NUMBER OF SEQ ID NOS: 2276  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 762  
; LENGTH: 6158  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; OTHER INFORMATION: n=a,t,g or c  
US-09-954-456-762

Query Match 8.4%; Score 239.8; DB 1C; Length 6158;  
Best Local Similarity 50.8%; Pred. No. 5.4e-55;  
Matches 756; Conservative 0; Mismatches 687; Indels 44; Gaps 6;

Qy 1326 TTGTCTCTCGGGAAACACAGGACTTCAGAGGCCCAAGAGTGCACCTGTGACTGCTCTGGAA 1385  
Db 3266 TGGAGCTCAGGCTGACCTGGACTGAAAGAGGGGAAGTCCCCAGGGCCACACAGGTCC 3325  
Qy 1386 CCTTGGCTACCTTGACAACTGGTCAAGATGTAAGCTTGATATCAAGGAAATGCAAG 1445  
Db 3326 AGTTGGCTCACCAGAGAACCTGGGTACAGAGGTACAGCTGGCCCAATTGGTTACAGAG 3385  
Qy 1446 GACACACAGTGTTCAGGATCTCCAGGAATACAAAGGAGCTCGAGGACTACACAGATTACA 1505  
Db 3386 GGCCTCGGGAACCTCAGGTCCTCTCTGGTCCAGCTGAGAGAAAGTGTCTCTGGAGAAA 3445  
Qy 1506 AGGAGAACACAGGGCAGATGGTGACAAAGGTGATGTGGACTTCTTGTTTCTCTGGGCT 1565  
Db 3446 AGGTCCCAAGGGCTCGAGGAGAGATGGAGTCAAGGTCTCTGTGTCTCCAGGGCC 3505  
Qy 1566 CATGGCATGCCAGATCAAGAGGTGAATGGTGCACAAAGGAGACAAAGATCACCTGG 1625  
Db 3506 AGCTGGTCTCTCGGCTCCCTGGGAAAGACGAGACAAAGGTCAAAATGGTGAGCCGG 3565  
Qy 1626 ATTTTATGGCAAAAGGGTGCAAAAGGTGAAAAGGGAAATGCTGGCTTCCCTGCTGCC 1685  
Db 3566 ACAAAAGCGCAGAGGGTGGCAGGGAGAAATGGCCCTCCCGTCCCCAGGTCTTCA 3625  
Qy 1686 TGGACTGCTGGAGAACACAGGAACATGGAAGATGATTAATGGGTAGTCCCGGTTT 1745  
Db 3626 AGGACCAAGTTGGTGGCCCTCGAATTTGCTGGAGGTGATGGTGAACCAAGTCTCTAGAGACA 3685  
Qy 1746 CAAGGAGAACGAGGATCCCTGGTGTCTCCGGGACAGATGGAAACAGGGGAGAGCCTGG 1805  
Db 3686 GCAGGGGATGTTTGGCAAAAGGTGATGAGGTGCCAGAGGCTTCCCTGGACCTCTCCGG 3745  
Qy 1806 AATCGAGAGTTTCTGAAACCCAGGATTAATGGGCCAAAGAGGAGAAAT----- 1857  
Db 3746 TCCAAATAGTCTTCAAGGTCTGCCAGGCCACCTGGTGAAAGAGGTGAATATGGGATGT 3805  
Qy 1858 -----GGGCTCCAGACAGCAGGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1907  
Db 3806 TGTTCATGGGGGCCACCTGGTCTCCAGGGCCCAAGAGGCCCTCAAGTCCCAATGGAGC 3865  
Qy 1908 AATGGAAAGCAATGCTCACCAGGCCACCTGGGAACACCGGATCTAAGGGAACCAAGG 1967  
Db 3866 TGATGACCAACAGGACCCCAAGTTCGTGTGGTCACTTGGTGTGTGGAGAAAGGG 3925  
Qy 1968 TGAACCTGGAATTCAGGGATCCCTGGGGCTTCTGGGCTCAAGGGAGAACACAGAGCAAC 2027  
Db 3926 TGAACCTGGAGAACGAGGAAACCCAGGGCCCTCTGGGGAAGCAGGTGTAGGCGTCCCAA 3985  
Qy 2028 GGGTTCCCGCAGGAGACACAGGATACATGGGTTTACCGGGATTCAAGGAAAGAGGGGA 2087  
Db 3986 AGGAGAAAGAGAGAGAAAGGGAGAGCTGGTCCACCTGGAGCTGCTGGACCTCCAGGTGC 4045  
Qy 2089 CAAGGAAATCAAGGTGAAAAGGTATTTCAGGGTCAAAAGGGAGAGAAATGGAAACAGGG 2147  
Db 4046 CAAGGGCGCCAGGTGATGATGGCCCTAAGGGTAACCCGGGTCTCTGTGTCTCTGG 4105  
Qy 2148 AATTCAGGCAACAGGAAATCAAGGCATCATGGTGCATAAGAGAGAGAGAGAGAGAGAG 2207  
Db 4106 AGATCTGTCTCTGGGAACTTGGCCCTGCAGGTCAAGATGTGTGTGTGTGTGACAA 4165  
Qy 2208 GGGAGAACTGTGTCTCGAGGTGCCATTGGATCAAAAGGAGAAATCTGGGTGGATGGCTT 2267  
Db 4166 GGGTGAATGAGATCTCTGTCAACCGGGTCTCTGGGCCATCTGGTGGGTGGCC 4225  
Qy 2268 GATGGGCGCCAGGTCTCTAAGGGGCAACCTGGGGATCCAGGTCTCTCAGGAGACCCCGAG 2327  
Db 4226 ACCAGTCTCTTGAAAAACAGGTCTCTCTGGAGTGCAGGTGCAGGAGGAGAGAGAGAG 4285  
Qy 2328 TTGTATGGGAAGCCCGGAGAGAGTTTTCAGAACAAATTTATTCGAACAAGTTTCCACAG- 2386  
Db 4286 TGAATAAGGTCTTAAGGGGGAAGCAGGTGCAAGAGGTCTCTCTGGAAAGACCCGCCAGT 4345  
Qy 2387 ATGTAAATAGAGCCCGAGCTACCACTCTTACTTCAGAGTGGNAGAAATAGAAATTTGTGATC 2446

Db 4263 AGATCCTGGTCTCTCTGGGAACCTTGGCCCTGACGCTCAAGATGGTGTGGTGTGACAA 4322  
QY 2208 GGGAGAACCTGGTCTCGAGGTGCCATTGGATCAAAAGGAAATCTGGGTGGATGCTT 2267  
Db 4323 GGGTGAAGATGAGATCCTGTGTCAACCGGGTCCCTCTGGCCCATCTGTGAGGTGGCCCC 4382  
QY 2268 GATGGGCCCCGAGCTCTAAGGGGCAACCTGGGATCCAGGTCTCAGGGACCCCCAGG 2327  
Db 4383 ACCAGTCTCTCTGGAAAACGAGGTCTCTGGAGTGCAGTGCAGAGGGAACAAG 4442  
QY 2328 TTGGAATGGGAAGCCCGGAAGAGATTTTCAGAACAAATTTATTCGACAAAGTTTCAAG- 2386  
Db 4443 TGAATAAGGTGCTAAGGGGGAAGCAGGTGCGAAGAGTCTCTCTGGAAAAACCGGCCAGT 4502  
QY 2387 ATGTAATAAGACCCAGCTACAGCTTACTTTCAGATGGAAGATTAGAATTTGTGTC 2446  
Db 4503 CGTCTCAGGACCTGAGGAAACCTGGTCCAGA-----AGGTCTTCGGGGC 4551  
QY 2447 ATTGCTCTGCCAATCGCTCCCGGGTATTCTTGGGCCACCTGGTCCGATAGGCCCAG 2506  
Db 4552 ATCCCTGGTCC--TGTGGAGAAACAAGTCTCCCTGGAGCTGCAGGCCAAGATGACCCAC 4609  
QY 2507 AGGTCCAGAGGATTAAGTCTGTTGGCAGGAGATGTTCTCTGGATTTAGTGGGTG 2566  
Db 4610 CTGGTCTTATGGGACCTCTGGCTTACCTGGTCTCAAGGTGACCTGGCTCCCAAGGGTG 4669  
QY 2567 TCCCTGGACGCTCCAGGTGTCAAGAGATTAAGAGGCTTACAGGAGAAATGGGGAAAAAG 2626  
Db 4670 AAAAGGACATCTCTGTTTAAATGGCTGATTGGTCTCCAGGAGACAAGGGGAAAAAG 4729  
QY 2627 GGAGCCAAAGGTTTGGGTATCTCTGGAGAACAGGTCTCTCTGGTCCCTCCAGGTCCAGAGG 2686  
Db 4730 GTGACCGAGGGCTC-----CCTGGAATCAAGGATCTCCAGGAGCAAAAGGGATGGG 4783  
QY 2687 GCCCTCTGGAATAGCAAGAGGTCTCCAGGAGACCCAGGTCTCCCTGGCAAGATG 2746  
Db 4784 GAATTCCTGGTCT-----GCTGTCTCTTATGTTCCACCTGGTCTCTCAGGCTTACAG 4837  
QY 2747 GAGACCATGGAACACTGGAATCCAAGGCAACAGGCCCCCCCAGGC 2793  
Db 4838 GTCTCTCAAGGCCCAAGGGTACAAAGGCTCTACTGGACCCGCTGGC 4884

## RESULT 14

US-09-919-497-6  
; Sequence 6, Application US/09919497  
; Patent No. US20020106662A1  
; GENERAL INFORMATION:  
; APPLICANT: Muller, George L.  
; TITLE OF INVENTION: PROGNOSTIC CLASSIFICATION OF ENDOMETRIAL CANCER  
; FILE REFERENCE: B0801/7225  
; CURRENT APPLICATION NUMBER: US/09/919,497  
; CURRENT FILING DATE: 2001-07-31  
; PRIOR APPLICATION NUMBER: US 60/221,735  
; PRIOR FILING DATE: 2000-07-31  
; NUMBER OF SEQ ID NOS: 100  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 6  
; LENGTH: 6158  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: Unsure  
; LOCATION: (2434)..(2434)  
; OTHER INFORMATION: n = a, c, g or t/u

Query Match 8.4%; Score 239.8; DB 10; Length 6158;  
Best Local Similarity 50.8%; Pred. No. 5.4e-55;  
Matches 756; Conservative 0; Mismatches 687; Indels 44; Gaps 6;

QY 1326 TTGTCTCTCGGAAAAACAGGACTTCAAGGCCCCCAAGGTGACCTGTGACTCGCTGGGAA 1395

Db 3266 TGAGCTCAGGCTGACCTGACTGAAGAGGAGGGAAAGTCCCCAGGGCCCCACAGGTCC 3325  
QY 1386 CCTTGGTACCTTGGACACCTTGGTCAAGATGGTAAGCCTGGATATCAGGGAATTTGCAGG 1445  
Db 3326 AGTTGGCTCACCAGGAGAACTGGGTACAGCTGGCCCAATTTGGTTTACGAGG 3385  
QY 1446 GACACCAGTGTTCAGGATCTCCAGGAATACAAGGAGCTCCAGGACTACCAAGTTACAA 1505  
Db 3386 GGGCCCCGACCTCAGGTCTCTCTGGTCCAGCTGGAGAGAAAGTGTCTCTTGGAGAAA 3445  
QY 1506 AGGAGAACCCAGGCCAGATGATGTAACAAGGTGATCTGTGACTTCTGTGTTTCTTGGGT 1565  
Db 3446 AGTCTCCCAAGGCTCTGACAGGAGAGATGGAGTTCAAGGTCTGTGTCTCCAGGGGCC 3505  
QY 1566 TCATGSCATGCCAGGATCAAGGGTGAATGGTCCCAAGGAGACAAGGATCACTCTGG 1625  
Db 3506 AGCTGGTCTCTGGCTCCCTCTGGGGAAGACGAGACAAGGGTGAATTTGGTGAAGCCGGG 3565  
QY 1626 ATTTTATGGCAAAAAGGTGCAAAAAGGTGAAAAGGGGAATGCTGGCTTCTCTTGGCTCTCC 1685  
Db 3566 ACAAAAAGGACGCAAGGGTGGCAAGGAGAAAATGSCCTCCCTCCCAAGGTCTTCA 3625  
QY 1686 TGGACCTGTGAGAACCCAGGAGACATGGAAGATGGAATTAATGGGTAGTCCCGTTT 1745  
Db 3626 AGGACAGTTGTGCTCCCTCTGAAATTCCTGGAGGTGATGAGTCCCTAGAGGACA 3685  
QY 1746 CAAGGAGAAAGCAGGATCCCTGGTCTCCGGGCGAGGATGGAACACCGGAGAGCCTGG 1805  
Db 3686 CGAGGGATGTTTGGCAAAAAGGTGATGAGGTGCCAGAGCTTCCCTGGACCTCTCTGG 3745  
QY 1806 AATCCAGGATTTCTCTGAAACCGAGGATTAATGGGCCAAAAAGGAGGAAAT----- 1857  
Db 3746 TCAATAGTCTTTCAGGGTCTCCAGGCCCACTGGTGAAAGGCTGAAATATGGGATGT 3805  
QY 1859 -----GGCCTCCAGGACAGCAAGGAAAAAGAGCCCGAGGATGCTCTGGTT 1907  
Db 3806 TGGTCTCTGGGGCCACCTGCTCCAGGCCCAAGAGGCCCTCAAGGTCCCAATGGAGC 3865  
QY 1908 AATGGGAAGCAATGGCTCACAGGCCAGCTCGAACACCGGATCTAAGGGAAGCAAGG 1967  
Db 3866 TGAATGACCAAGGACCCCTCAGTCTCTGTGTTAGTGGTGTGGAGAAAGG 3925  
QY 1968 TGAACCTGGAATTCAGGATGCTCTGGGCTTCTGGGCTCAAGGAGAACCCAGGAGCAAC 2027  
Db 3926 TGAACCTGAGAGCAGGAAACCCAGGGCTCTCTGGGAAGCAGGTGAGGCGTCCCAA 3985  
QY 2028 GGTTCCTCCAGGAAACCCAGGATCATGSGTTTACCCGGATTCAGGNAAAAAGGGGA 2087  
Db 3986 AGAGAAAGAGGAGAGAAAGGGGAGCTGCTCCAGCTGTGAGCTGTGAGCTCCAGGTGC 4045  
QY 2088 CAAAGAAAATCAAGGTGAAAAAGGTATTCAAGGTCAAAAGGAGAAAAATGGAAGACAGG 2147  
Db 4046 CAAAGGGCCGCGNGGTGATGATGGCCCTAAGGTAAACCCGGTCTCTGTGTTTCTCTGG 4105  
QY 2148 AATTCCAGGGCAACAGGAAATTCAGGCAATCATGTGCAAAAGGAGAGAGGTGAATA 2207  
Db 4106 AGATCTCTCTCTCTGGGAACTTGGGCCCTGCAAGTCAAGATGTTGTGTGTGACAA 4165  
QY 2208 GGGAGAACCTGTGTCCAGGTGCCATTGGATCAAAAGGAGAACTCTGGGTGGATGGCTT 2267  
Db 4166 GGTGAAGATGGAGATCTGTGTCAACCGGCTCTCTGGCCATCTGTGAGCTGGGCC 4225  
QY 2268 GATGGGGCCCGAGGTCTTAAGGGGCAACCTGGGGATCCAGGTCTCAGGACCCCCAGG 2327  
Db 4226 ACCAGGTCTCTCTGGAAAAACAGGTCTCTCTGTGAGTGCAGGTGCAGAGGGAAGCAAGG 4285  
QY 2328 TTGATGGGAAGCCCGGAGAGAGTTCAGAACAAATTTATTCGACAAAGTTTGCACAG- 2386  
Db 4286 TGAATAAGGTGCTAAGGGGGAAGCAGGTGCAAGAGGTCTCTCTGGAAAAACCGGCCAGT 4345  
QY 2387 ATGTAATAGACCCAGCTACAGTCTTACTTACAGTGGGAAGAAATTAGAAATTTGTGATC 2446

```

QY 1928 CAGCCAGCCTGGACACCGGATCTAAGGAGCAAGGATGAACCTGGAAATCAAGGGA 1987
Db 1271 CTGGCCCTCTGGGATTAATGTAGTCTCTGTGTAAAGCGGAAATGGGTCCCGCTGGCA 1330
QY 1988 TGCCTGGGCTTCTGGGCTCAAGGAGAAACAGGAGCAACGGGTTCCTCCAGGAGAACCCAG 2047
Db 1331 TTCTGGAGCTCTGGATCTGATGGAGCCCGGGTCTCTCCAGGACCAAGCTGGTGTAAATG 1390
QY 2048 GATACATGGGTTTACCCGGGATTCAGGAAAAAAGGGGACAAAGGAATCAAGGTGA 2107
Db 1391 GTGCTCTGGATCTCGAGGTGTCTGAGTCTGTAAGAAATGGTGCCAAAGGAGAGC 1450
QY 2108 AAGGTATTACGGGTCAAAAGGAGAAATGGAAGACAGGAAATTCAGGCAACAGGAA 2167
Db 1451 CCGGACACCTGGTCAACCGGCTGAGCTGTATTCCAGGTGTCTCCAGGAGCTAAGGCG 1510
QY 2168 TTCAGGCAATCATGGTGCAAAAGGAGAGAGAGTGAAAGGAGAACTGTGTGTCCGAG 2227
Db 1511 AAGATGCAAGGATGATCACTGAGAACTGTGCAAAATGGGCTTCCAGGAGCTGCAG 1570
QY 2228 GTGCCATTGGATCAAAAGGAGAAATCTGGGTGGATGCTTGTATGGGCGCCGAGGTCCTA 2287
Db 1571 GAGAAAGGGTGCCCTGGGTTCGAGAACCTGTCTGACCAATGGATCCCAGGAGAA 1630
QY 2288 AGGGGCAACCTGGGATCCAGGTCTCTCAGGACCCAGGTTTGGATGGGAAACCCGGA 2347
Db 1631 AGGGTCTGTCTGGAGAGCTGTCTCCAGGCCCTGACAGGCCCCAGAGGAGCTCTGGAG 1690
QY 2348 GAGAGTTTTCAGAACATTTATTTCAGACAGTTTCACAGATGTAAAGAGCCAGCTAC 2407
Db 1691 AACCTGGCAGAGATGGGCTCCCTGGAGTCCAGGAATGAGGGGATGC---CCGGAAGTC 1747
QY 2408 CAGTCTTACTTCAGATGGAAGAAATAGAAATTTGATCATTTGCTCTCCCAATAGGCT 2467
Db 1748 CAGGAGGACAGGAGTATGAGGAAACAGGCGCTCCCGGAGTCAAGGAGAAATGGTC 1807
QY 2468 CCCGGGTATTCCTGGGCACTCTGTCTGATAGGAGGAGGAGGAGGAGGAGGAGGAGG 2527
Db 1808 GACCAAGTCTCTGGGCACTCTGTCTCCCGAGGTCAGCTGTGTCTATGGGCTTCCCG 1867
QY 2528 GTTTCAGGAGAGAGTGTCTCTCGATTAGTGGGTCTCCCTGGAGCTCCAGGTGTCA 2587
Db 1868 GTCTTAAGGAAATATGTGTCTCTGTAAAGATGAGAACAGGTTGGCCCTCGAGAGAC 1927
QY 2588 GAGGATTAAGAGGCTTACAGGAGAAATGGGAGAAAGGAGGAGCAAGGTTTGGTATC 2647
Db 1928 CTGGCCCTCAGGGTCTCTGGAAGAAATGGTGAATCTGGACCTCAAGGACCCCGAGGC 1987
QY 2648 CTGGAGAACAGGCTCTCTGGTCCCCAGGTCAGAGGCGCTCTCTGGAAATAGCAAG 2707
Db 1988 CTACTGGGCTGG---TGGTCAAAAGGAGACACAGGACCCCTCTGGTCCCAAGGATTAC 2044
QY 2708 AAGGTCTTCCAGGAGACCCAGGTCTCTCTGGCAAGATGGAGACCATCGAAACCTGGAA 2767
Db 2045 AAGGCTTCTGGTACAGGTGTCTCTCAGGAGAAATGGAAACCTGGGNAACAGGTC 2104
QY 2768 TCCAAAGGCAACAGGCCCCCAGG 2792
Db 2105 CAAAGGGTATGCCCGTGACCTGG 2129

```

RESULT 13  
 US-10-096-534-15  
 ; Sequence 15, Application US/10096534  
 ; Publication No. US2003016887A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: The Brigham and Women's Hospital, Inc.  
 ; APPLICANT: Yates, Karen  
 ; APPLICANT: Mizuno, Shuichi  
 ; APPLICANT: Glowacki, Julie  
 ; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF SKELETAL DEGENERATION CONDITIONS  
 ; FILE REFERENCE: B0801/7244/KA/ERP

```

; CURRENT APPLICATION NUMBER: US/10/096,534
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: JS 60/274,980
; PRIOR FILING DATE: 2001-03-12
; NUMBER OF SEQ ID NOS: 79
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 15
; LENGTH: 6319
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-096-534-15

Query Match      8.4%; Score 241.4; DB 12; Length 6319;
Rest local Similarity 50.9%; Pred. No. 2e-55;
Matches 757; Conservative 0; Mismatches 686; Indels 44; Gaps 6;

QY 1326 TTGTCTCTCGGAAAAACGAGGACTTCAAGGCCCCCAAGGTGACCTGGACTGCTGGGAA 1385
Db 3423 TCGAGCTCAGGCTCAGCTGGACCTGAAAGAGGGGAAAGGTCCCGGGGCCACAGGTCC 3482
QY 1386 CCCTGGCTTACCTGCAACACCTGCTCAAGATGGTAAGCCTGATATCAGGGAATTCAGG 1445
Db 3483 AGTTGGCTCACGAGAGAACGTGGGTGAGCAGGTACAGCTGSCCCCAATTGGTTTACCAGG 3542
QY 1446 GACACCAAGTGTTCAGGATCTCCAGGAATACAGGAGCTTCGAGGACTACAGGTTACAA 1505
Db 3543 CGGCCCGGACCTCAGGCTCTCTCTGCTCCAGCTGGAGAGAAAGGTCTCTCTGGAGAAA 3602
QY 1506 AGGAGAACCGGGGAGATGTGACNAGGGTGAATCGTGGACTTCTCTGGTTTCTCTGGCT 1565
Db 3603 AGTCCCCAAGGGCTCTCAGGAGAGATGAGTTTCAAGGTCTCTGTTGGTCTCCAGGGCC 3662
QY 1566 TCATGGCATGCCAGSATCAAAAGGTTGAAATGGTGTCCAAAGAGAGACAAAGGATCACCTGG 1625
Db 3663 AGCTGTCTCTCCGGCTCCCTGCGAAGACGAGACAAAGGTGAATTTGGTAGCCGG 3722
QY 1626 ATTTATGGCAAAAGGTCGAAAGCTGAAAGGGGAAATGCTGGCTTCCCTGGCTCCC 1685
Db 3723 ACNAAAGGAGCAGGAGGTGCAAGGGGAGAAATGGCCCTCCCGGTCCCCAGGTCTTCA 3782
QY 1686 TGGACCTCTCGAACAACGAGGACATGGAAGGATGGAAATGATGGGTAGTCCCGGTTT 1745
Db 3783 AGGACCATGTCCTGCTGCTGCTGAAATGCTGGAGGTGATGCTGACCAAGTCTTAGAGACA 3842
QY 1746 CAAGGAGAAACAGATCCCTGCTGCTCCGGGAGAGATGGAACACGCGGAGAGCCTGG 1805
Db 3843 GCAGGGGATGTTTGGGCAAAAGGTGATGAGGTTSCCAGAGGCTTCCCTGGACCTCTCTGG 3902
QY 1806 AATCCCAAGATTTCTCTGAAACCCAGGATTAATGGGCCAAAGGGAGAAATTT----- 1857
Db 3903 TCCATAGTCTTTCAGGGTCTGCGAGGCCCCCTGCTGTAAGGAGTGAATGGGGATGT 3962
QY 1858 -----GGGCTCTCAGACAGCAAGGAAAAAGAGCCCGAGGATGCTGCTGGTTT 1907
Db 3963 TGTTCATGCGGGGCCACCTGCTCTCCAGGCCCAAGAGGCCCTCAAGTCCCAATGAGC 4022
QY 1908 AATGGAGCAATGCTCAACAGGCCAGCTTGAACACCCGGGATCTAAGGAGCAAGG 1967
Db 4023 TGATGGACCAAGAGACCCCGAGTTCTGTGTTGTTCTAGTTGGTGTGAGAGAAAGGG 4082
QY 1968 TGAACCTGGAATTCAGAGGATGCTCGGGCTTCTGGGCTCAAGGGGAGAACCAAGAGCAAC 2027
Db 4083 TCACTGGAAGCAAGCAAAACCCAGGGCTCTCTGGGAGAGCAGGTGTAGCGGTCCCAA 4142
QY 2028 GGGTTCCCGAGAGAACCAAGATACATGGTTTACCCTGGGATTCAGGAAAAAGGAGG 2087
Db 4143 AGGAGAAAGAGAGAGAAAGGGAGAGTGTGCTGAGCTGAGCTGCTGAGCTCCAGGTGC 4202
QY 2088 CAAAGGAATCAAGGTGAAAGGATTTACAGGTCAAAAGGAGAGAAATGGAAGACAGG 2147
Db 4203 CAAGGGCGCGCAGGTGATGATGCTTAAAGGGTAACCCGGGCTCTGTTGTTCTCTGG 4262
QY 2148 AATTCAGGGCAACAGGAAATTCAGGCGCATTCATGGTGCAAAAGGAGAGAGAGTGA 2207

```

QY 1688 GACCTGCTGGAGAACCCAGGAGACATGGAAGGATGATTAATGGGTAGTCCCGTTTCA 1747  
 Db 1025 GGGCTCTGGTGGAGGAGCGCCAGGACTTCTTGGGGCTGCAGGTGCTCGGGGTAAATG 1084  
 QY 1748 AGGAGAAAGCAGGATCCCTGGTCTCCGGGCGAGATGGAAACACGGGGAGAGCCTGGAA 1807  
 Db 1085 ACGGTGCTCGAGCAGTATGCTCAACAGGCCCTCTCTGGTCTCTTGGAACTCCGGAT 1144  
 QY 1808 TCCAGGATTTCTCGAAACCGAGATTAATGGGCCAAAGGAGAGAAATTTGGGCTCCAG 1867  
 Db 1145 TCCCTGGATCCCTGGTCTAAGCGTGAATGGACCTCGAGGGTCTCTGGTTCAAATG 1204  
 QY 1868 GACAGCAAGGAAAAAGAGCCCGAGGATCCCTGGTTTAAATGGGAAGCAATGGCTCAC 1927  
 Db 1205 GTGCCCTGGACAAAGAGGAGAACTTGAACCTCAGGACACGCTGGTCTCAAGTCTTC 1264  
 QY 1928 CAGGCAGCCTCGAACACCGGATCTAAGGAGAGCAAAAGGTGAACCTTGGAAATTCAGGGA 1987  
 Db 1265 CTGGCCCTCTCTGGATTAATGTAGTCTCTGGTGTAAAGGGGAAATGGGTCCCGCTGCA 1324  
 QY 1988 TGCCTGGGCTTCTGGCTCAAGGAGAGAACCCAGGAGCAACGGGTTCCTCCAGGAGAACCG 2047  
 Db 1325 TTCTTGGACTCTGACTGATGGAGCCCGGGTCTCTCAGGACCAGCCGGTCTAATG 1384  
 QY 2048 GATACATGGTTTACCCCGGATTTCAAGGAAAAAGGGGACAAAGGAAATCAAGGTGAAA 2107  
 Db 1385 GTGCTCTGGACTGCGAGTGTGCGAGGTGAGCCCTGTGAAGATGCTGCCAAGAGAGC 1444  
 QY 2108 AAGGTATTGAGGTCAAAAGGAGAGAAATGGAAGACAGGGAATCCAGGCAACAGGAA 2167  
 Db 1445 CCGGACCAAGTGTGAACCGGTGAGGTGTGATTTCCAGGTGTTCAGGAGCTAAAGCG 1504  
 QY 2168 TTCAAGGGCCATCATGGTGCMAAGAGAGAGAGGTGAAAGGGAGAACCTGGTCTCCGAG 2227  
 Db 1505 AAGTGGCAAGATGATCACTGTGAGAACCTGTGCAAAAGGGCTTCAGAGAGTGCAG 1564  
 QY 2228 GTGCCATTGGATCAAAAGAGAAATCTGGGGTGGATGGCTTGAATGGGGCCCGAGGTCTTA 2287  
 Db 1565 GAGAAAGGGTGTCCCTGGTTCAGGAGACCTGTGTGACCAAAATGGCATCCAGAGAGAA 1624  
 QY 2288 AGGGCAACCTGGGATCCAGTCTCAGGAGACCCAGGTTTGGATGGGAAGCCCGAA 2347  
 Db 1625 AAGGTCTCTGGAGAGCGTGTGTCTCAGGCGCTTCAGGCGCCAGAGAGAGTGTGGAG 1684  
 QY 2348 GAGAGTTTTCAGACAAATTTATTCGACAAATTTGACACAGATGTATTAAGAGCCAGCTAC 2407  
 Db 1685 AACCTGGCAGAGATGCGCTCCCTGGAGTCCAGGAATGAGGGCATGCGTGCATGTC 1741  
 QY 2408 CAGTCTTACTTACAGTGTGAAGAAATAGAAATTTGATCATTTGCTGTCCCAACATGCT 2467  
 Db 1742 CAGGAGGACAGAGATGATGGAAACCAAGGCGCTCCCGGAAGTCAAGGAGAAATGCTC 1801  
 QY 2468 CCGCGGTATTCTTGGCCACCTGGTCTCCATAGGCCCGAGAGGTCCTCCAGAGATTTACCTG 2527  
 Db 1802 GACCAAGTCTCTGGGCCATCTGTCTCCGAGGTGAGCTGCTGTGATGGCTTCCCG 1861  
 QY 2528 GTTTCAGGAGAGATGTGTCTTGGATTAATGGGTGCTCCTGGAGCTCCAGGTGCTCA 2587  
 Db 1862 GTCTTAAAGGAAATGATGTGTCTCTGGTGAAGATGGAGAACAGGTGGCCCTGAGGAC 1921  
 QY 2588 GAGGATTTAAAGGCTTACAGGAAGAAATGGGGAAGAAAGGAGCCAAAGGTTTGGGTATC 2647  
 Db 1922 CTGGCCCTCAGGGTCTCTCTGGAAGAAATGGTGAATCTGAACTCAGGACCCCGAGG 1981  
 QY 2648 CTGGAACAAAGTCTCTCTGTGTCCTCCAGGTTCAGAGGGCCCTCTCTGGAATTAAGCAAG 2707  
 Db 1982 CTACTGGGCTGGTCTGTTGAGAGAAATGGTGAATCTGAACTCAGGACCCCGAGG 2038  
 QY 2708 AAGGTCTCCAGAGACCCAGGTCTCTCTGGCAAGATGGAGACCACTGGAAACCTGGAA 2767  
 Db 2039 AAGGCTTGGCTGTACAGGTGGTCTCTCAGGAGAAATGGAAATCTGGGAAACAGGTC 2098

QY 2768 TCCAAGGGCAACCCAGGCCCCCAGG 2792  
 Db 2099 CAAAGGGTGAATGCCGTGCACCTGG 2123  
 RESULT 12  
 US-10-198-646-9949  
 ; Sequence 9949, Application JS(10-98846  
 ; Publication No. US2003009974A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lillie, James  
 ; APPLICANT: Xu, Yongyao  
 ; APPLICANT: Wang, Youzhen  
 ; APPLICANT: Steinmann, Kathleen  
 ; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS  
 ; FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
 ; THERAPY OF BREAST CANCER  
 ; FILE REFERENCE: MRI-049  
 ; CURRENT APPLICATION NUMBER: US/10-198,846  
 ; PRIOR FILING DATE: 2002-07-18  
 ; PRIOR APPLICATION NUMBER: 60/306,220  
 ; NUMBER OF SEQ ID NOS: 14084  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 9949  
 ; LENGTH: 8664  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-10-198-646-9949  
 Query Match: 8.5%; Score 242.6; DB 14; Length 8564;  
 Best Local Similarity 48.8%; Pred. No. 1.2e-55;  
 Matches 715; Conservative 0; Mismatches 744; Indels 6; Gaps 2;  
 QY 1328 GTCCCTCCGGAAGAACCCAGGACTTCAAGGCCCAAGGTGACCCCTGGACTGCTGGAAACC 1387  
 Db 671 GTCCCCCTGGTACATCTGTGTCATCTCTGGTTCCCTTGGATCTCCAGGATACCAAGACCCC 730  
 QY 1388 CTGGCTACCTCGAAACCTGCTCAAGATGTGAACCTGGATATCAGGGAATTCAGGGA 1447  
 Db 731 CTGTTGAACCTGGCAAGCTGTCTTCAGGCCCTCCAGGACCTCTGTGTCTATAGGTC 790  
 QY 1448 CACCAAGTCTTCAGGACTCTCCAGGAATACAGGAGCTCGAGGACTACAGGTACAAAG 1507  
 Db 791 CATCTGTCTCTGTGAAAGATGGAGAAATAGGTAGACCCGACGACTGAGAGCCAG 850  
 QY 1508 GAGAACCCAGGGCGAGATGGTGAACAGGGTGAATCTGTGGTCTTCTGTGCTTCTGTGGCTTC 1567  
 Db 851 GATTGCTTGGACTCTCAGGTAACAAAGGTCCAGTGGGATACCTGGATTCCCTGGTATGA 910  
 QY 1568 ATGGCATGCCAGATCAAGGGTGAATGGGTGCCAAAGGAGACAAAGATCACCCTGGAT 1627  
 Db 911 AAGGACACAGAGGCTTCGATGGACGAAATGGAGAAAGGGTGAACAGGTGCTCTCTGGAT 970  
 QY 1628 TTATATGGCAAAAGGGTGCAAAAGGTGAAAAGGGGAATGCTGGCTTCCCTTGGCTTCCCTG 1687  
 Db 971 TAAAGGGTGAATGGTCTTCAGGCGAAATGGAGCTCTCGACCCATGGTCCAGAG 1030  
 QY 1688 GACCTGCTGAGAAACAGGAGACATGGAAGATGGAATTAATGGGTAGTCCCGTTTCA 1747  
 Db 1031 GGGCTCTCTGTGAGCAGGACCGCCAGGACTTCTCTGGGGCTGCAGAGTCTCGGGGTAATG 1090  
 QY 1748 AGGGAAGACAGATCCCTGTGTCTCCGGGCGAGGATGGAAACCGGGAGAGCTCGAA 1807  
 Db 1091 ACGGTGCTCGAGGAGTGAATGGTCAACAGGCGCTTCTGTGCTCTGGAATCTGCCGAT 1150  
 QY 1808 TCCAGGATTTCTCGAAACCGAGATTAATGGGCCAAAGGAGAAATTTGGGCTCCAG 1867  
 Db 1151 TCCCTGGATCCCTGTGTCTAAGGGTGAATTTGGACCTCGAGGGTCTCTCTGTCAAATG 1210  
 QY 1868 GACAGCAAGGAAAAAGGAGCCCGAGGATGCTCTGGTTTAAATGGGAAGCAATGGCTCAC 1927  
 Db 1211 GTCCCTCTGCAAGAGAGGAAACCTGGACCTCAGGGACACGCTGTGTCTCAAGGTCTCTC 1270

```
Db 710 CTGGTGAACCTGGGCAAGCTGGTCTTTCAGGCCCTCCAGGACCTCTCTGGTGTATAGTTC 769
QY 1448 CACAGGTTGTTCCAGGATCTCCAGGAATACAGGAGCTCCAGGACTACCAAGTTACAAG 1507
Db 770 CATCTGGTCTCTGGAAAGATGGAAATCAGTATAGTACCCGACGACCTGGAGCGGAG 829
QY 1508 GAGAACCCAGGCGGAGATGGTGACAAAGGTTGATCCTGGACTTCTCTGGTTTCTCTGGGCTTC 1567
Db 830 GATTGCTGGACCTCCAGGTATCAAGGTCAGCTGGGATACCTGGATTCTCTGGTATGA 889
QY 1568 ATGGCATGCCAGGATCAAGGTTGAATGGTGCCAAAGAGAGACAAGGATCACTGGAT 1627
Db 890 AAGGACACAGAGGCTTCGATGACGAATGGAGAAAGGGTGAACAGGTGCTCTCTGGAT 949
QY 1628 TTTATGGCAAAAGGGTGCAAAAGGTGAAAAGGGAATGCTGGCTTCTCTGGCTCCTCTG 1687
Db 950 TAAAGGTGAATATGGTCTCCAGCGGAATATGGAGCTCTCTGACCATGGTCCAAAG 1009
QY 1688 GACCTGTGGAGAACCCAGGAAGACATGGAAGAGATGGAATTAATGGGTAGTCCCGGTTTCA 1747
Db 1010 GGGCTCTGGTGAGCGAGGACCGCCAGGACTTCTCTGGGGCTGCAGGTGCTCGGGTAATG 1069
QY 1748 AGGAGAGACAGGATCCCTGGTCTCCGGGAGGATGGAACACGGGAGAGCCTGGAA 1807
Db 1070 ACGGTGCTCGAGGAGTATGATGATCAACAGGACCTCTCTGGTCTCTCGAACTGCCGGAT 1129
QY 1808 TCCAGGATTTCTCTGGAAACCGAGATTAATATGGGCCAAAGAGGAGAAATTTGGGCTCCAG 1867
Db 1130 TCCCTGGATCCCTGGTCTAAGGTTGAAGTTGACCTGCGAGGTCTCTGGTTCATATG 1189
QY 1868 CACAGCAAGGAAAGAAAGAGCCAGGATGCTCTGGTTTAATGGGAAGCAATGGCTCAC 1927
Db 1190 GTGCCCTCTGGCAAGAGAGGAACTCTGACCTCAGGACACGCTGCTCAAGGCTCCTC 1249
QY 1928 CAGGACAGCTGGAAACCGGATCTTAAGGAAGCAAGGTTGAACCTTGAATTTCAAGGGA 1987
Db 1250 CTGGCTCTCTGGGATTAATGGTAGTCTCTGGTGAAGGCAATGGTCCCGCTGCCA 1309
QY 1988 TGCTCTGGGCTTCTGGGCTCAAGGAGAACCAAGAGCAACCGGTTCCCGGAGGAACAG 2047
Db 1310 TTCTGTGAGCTCTCTGGATGATGGAGCCCGGGTCTCTCCAGGACCAAGCGGTCTAATG 1369
QY 2048 GATACATGGGTTTACCCGGATTCAGGAAAGAGGAGGCAAGAGGAATCAAGTCAAA 2107
Db 1370 GTGCTCTGGACTCGAGGTGGTGCAGGTGAGCCTGAGTGAAGATGGTGCCAAAGGAGGC 1429
QY 2108 AAGGTATTCAGGTTCAAAAGGAGGAGAAATGGAAGACAGGGAATTCAGGGCAACAGGGAA 2167
Db 1430 CCGACACAGGTGGTGAACGCGGTGAGCTGGTATTCAGGTGTTCCAGGAGCTAAAGGCG 1489
QY 2168 TTCAGGCGCATCATGGTGCAAAAGAGAGAGAGGTGAAAGGAGAGAACCTGGTCTCGAG 2227
Db 1490 AAGATGGCAAGATGGATCACTGGAGAACCTGGTGCAAAATGGGCTTCCAGGAGCTCGAG 1549
QY 2228 GTGCCATTTGATCAAAAGAGAACTCTGGGTGGATGGCTTGAATGGGCGCCGAGTCTCA 2287
Db 1550 GAGAAAGGGTGCCCTGGTTCCAGGAGACCTCTGGACCAATGGCATCCAGGAGAAA 1609
QY 2288 AGGGCAACCTGGGATCCAGGTCTCTAGGAGACCCCGAGTTTGGATGGGAAGCCCGGAA 2347
Db 1610 AGGGTCTCTGGAGAGCGTGGTGTCTCAGGCGCTGCGAGGCGCCAGAGGAGCTCTCGAG 1669
QY 2348 GAGAGTTTTCAGAACTTATTTCAGCAAGTTTGCACAGATTTGACAGATGTAATAGAGCCAGCTAC 2407
Db 1670 AACCTGGCAGATGGCGTCTCGAGTCCAGGATCCAGGAATGAGGGGATGC---CCGGAATG 1726
QY 2408 CAGTCTTACTTTCAGAGTGCAGAAATTAGAAATTTGATCATTTGCTCTGCCCAACATGGCT 2467
Db 1727 CAGGAGGACCAAGAGTATGGGAACCAAGGCGCTCCCGGAGTCAAGGAGAAATGGTCTC 1786
QY 2468 CCGCGGTTATCTCTGGGCACTCTGGTCCGATAGGCGCCAGAGGGTCCCAAGAGGATTACCTG 2527
```

```
Db 1787 GACCAGGTCTCTCTGGGCCATCTGGTCCCCAGAGTCAGCCTGGTGTGTCTATGGGTTCCCCG 1846
QY 2528 GTTTCGCCAAGAGATGGTGTCTCTGGATTAGTGGGTGTCTCTGGACGTCCAGGTGTCA 2587
Db 1847 GTCTTAAAGGAATGATGGTGTCTCTGGTAAATGGAGACGAGGTGGCCCTGGAGGC 1906
QY 2588 GAGGATTTAAAGGCTTACACAGGAAGAAATGGGAAAAAGGAGCCAAAGGTTTGGGTATC 2647
Db 1907 CTGGCCCTCAGGCTCTCTCTGGAAAGAAATGTGAAACTGGACCTCAAGGACCCCAAGGC 1966
QY 2648 CTGGAGAAACAAGGCTCTCTCTGGTCCCCCAAGTCAGAGGGCCCTCTCTGGAAATAGCAAG 2707
Db 1967 CTACTGGGCCCTGG---TGGTGACAAAGAGACACAGGACCCCTCTGGTCCCAAGGATTAC 2023
QY 2708 AAGGTCTCTCCAGGAGACCCAGGTCTCTCTGGCAAGATGGAGACCATGGAAAACTGGAA 2767
Db 2024 AAGGCTTCTCTGGTGTACAGGTGTCTCTCCAGGAGAAATGGAAAACTGGGGAACCAAGTC 2083
QY 2768 TCCAAAGGCAACCAAGGCCCCCGCCAGG 2792
Db 2084 CAAAGGCTCATGCCCGTGCACCTGG 2108

RESULT 11
US-10-096-534-13
; Sequence 13, Application US/10096534
; Publication No. US20030166887A1
; GENERAL INFORMATION:
; APPLICANT: The Brigham and Women's Hospital, Inc.
; APPLICANT: Yates, Karen
; APPLICANT: Mizuno, Shuichi
; APPLICANT: G.owacki, Julie
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF SKELETAL DEGENERATION CONDITIONS
; FILE REFERENCE: B08017244/KA/ERP
; CURRENT APPLICATION NUMBER: US/10/096,534
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: US 60/274,980
; PRIOR FILING DATE: 2001-03-12
; NUMBER OF SEQ ID NOS: 79
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 13
; LENGTH: 5489
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-096-534-13

Query Watch 8.5%; Score 242.6; DB 12; Length 5489;
Best Local Similarity 48.8%; Pred. No. 8.3e-56;
Matches 725; Conservative 0; Mismatches 744; Indels 6; Gaps 2;

QY 1328 GTCTCTCCGGGAAAACAGGACTTCAAGSCCCCAAGGTGACCTGGACTGCTCGGAACC 1387
Db 665 GTCCCCCTGGTACATCTGGTTCATCTGGTTCCTCTGGATCTCCAGGATACCAAGGACCCC 724
QY 1388 CTGGCTTACCTTGGACAACTGTCTAAGATGTTAAGCCTTGGATATCAGGAAATTCAGGGA 1447
Db 725 CTGGTGAACCTGGGCAAGCTGTCTCTCAGGCCCTCCAGGACCTCTCTGGTCTATAGTC 784
QY 1448 CACCAAGGTGTTCCAGGATCTCCAGGAATACAAAGAGCTCCAGGACTACCAAGTTACAAG 1507
Db 785 CATCTGGTCTCTCTGGTGTGAAAGATGGAATCAGGTAGACCCGGAACGACCTGGAGAGGAG 844
QY 1508 GAGAACCAAGGCGAGATGTCACAAAGGTGATCGTGGACTCTCTGGTTTCTCTGGGCTTC 1567
Db 845 GATTGCTCTGGACCTCCAGGTATCAAGGTCCAGTGGGATACCTGGATTCTCTGGTATGA 904
QY 1569 ATGGCATGCCAGGATCAAAAGGTGAAATGGGTGCCAAGGAGACAAGAGATCACTGGAT 1627
Db 909 AAGGACACAGAGGCTTCGATCGACGAATGAGAAAGGGTGAACAGGTGCTCTCTGGAT 964
QY 1628 TTTATGGCAAAAGGGTGCAAAAGGTGAAAAGGGAATGTGGTTCCTCTGGGCTCTCTG 1687
Db 965 TAAAGGTTGAAATGGTCTCTCCAGCGGAAATGAGGTCTCTGGACCCATCGGTCCAAAGAG 1024
```

Qy 1508 GAGAACGAGGCGAGATGCTGACAAAGGTGATCGTGGACTTCTCTGGTTTCTCTGGCTTC 1567  
 Db 830 GATTGCTCGACCTCCAGGTATCAAGAGTCCAGCTGGGATACCTGGATTCCCTGGTATGA 889  
 Qy 1568 ATGGCATGCCAGATCAAGGGTGAATGGTGGCCAAAGAGACAAGGATCACTGGAT 1627  
 Db 890 AAGGACACAGAGGCTTCGATGGAGCAATCGAGAAAGGGTGAACAGGCTGCTCTGGAT 949  
 Qy 1628 TTTATGCAAAAAGGGTGCAAAAGGTGAAAGGGGAATGCTGCTTCCCTGGGCTCCCTG 1687  
 Db 950 TAAAGGTGAATGGTCTCCAGGCGAAATGGAGCTCTGACCCATGGTCCAGAG 1009  
 Qy 1688 GACTGCTGAGAACACAGGAGACATGGAAGAGATGATTAATGGGTAGTCCCGTTTCA 1747  
 Db 1010 GGGCTCTGCTGAGCGAGGACGGCCAGGACTTCTCTGGGCTGCGAGTGTCTCGGGTAATG 1069  
 Qy 1748 AGGGAAGACAGGATCCCTCTGGTCTCCGGGGAGGATGGAACACGGGGAGAGCCCTGAA 1807  
 Db 1070 ACGTGTCTGAGGAGTGTGTTCAACGAGCCCTCTCTGTCTCTCTGAACTGCGGAT 1129  
 Qy 1808 TCCAGGATTTCTGGAAACCGAGGATTAATGGGCCAAAGAGGAGAAATGGGCCCTCCAG 1867  
 Db 1130 TCCCTGATCCCTGGTCTGAAGGGTGAAGTTGGACCTGCGAGGTCTCTCTGGTTCAAATG 1189  
 Qy 1868 GACAGCAAGAAAGAGAGGACCCAGGATGCTGTGTTAATGGGAGCAATGGCTCAC 1927  
 Db 1190 GTGCCCCCTGCAAGAGGAGAACCTCGACCTCAGGACACGCTGTGTCTCAAGGTCCTC 1249  
 Qy 1928 CAGGCCAGCTGGAAACCGGGATCTAAGGAGCAAGGATGAACCTGGAATCAAGGGA -987  
 Db 1250 CTGGCCCTCTGGATTAATGGTGTCTGTGTAAAGGCGAATGGGTCCGCTGGCA 1309  
 Qy 1988 TGCTGGGCTTCTGGCTCAAGGAGAAACAGGAGCAACGGGTTCCTCCAGGAGAACAG 2047  
 Db 1310 TTCTGAGCTCTGGACTGATGGAGCCCGGGTCTCTCCAGGACACGCGGTGCTAATG 1369  
 Qy 2048 GATACATGGGTTTACCGGATTCAGGAAAGAGGGGCAAGGGAATCAAGGTGAA 2107  
 Db 1370 GTGCTCTGACTCGAGGTGTGAGGTGAGCCTGTGTAGAAATGGTGCCTCAAGGAGAGC 1429  
 Qy 2108 AAGTATTCCAGGTCAAAAGGGAGAAATGGAAGACAGGGAATCCAGGGCAACAGGAA 2167  
 Db 1430 CCGACACAGTGTGAACGGGTGAGCTGTATTCAGGTGTTCAGGAGCTAAGGGC 1489  
 Qy 2168 TTCAAGGCAATCATGGTGCAAAAGGAGAGAGAGTGAAAGAGGAGAACCTGGTGTCCGAG 2227  
 Db 1490 AAGATGGCAAGGATGATCATCTGGAGAACCTGTGTGCAATGGGCTTCCAGGAGCTGAG 1549  
 Qy 2228 GTGCCATTGGATCAAAAGGAGAACTCTGGGTGGATGCTTGTATGGGSCCGCAGGTCTTA 2287  
 Db 1550 GAGAAAGGGTGCCCTGGGTTCGAGGACCTGTCTGGACCAATGGCATCCAGGAGAA 1609  
 Qy 2288 AGGGGCAACCTGGGATCCAGGTCTCTCAGGACCCCGAGGTTTGGATGGGAAGCCGGAA 2347  
 Db 1610 AGGGTCTCTGGAGAGCGTGTCTCCAGGCCCTGAGGCGCCAGAGGAGCTGCTGGAG 1669  
 Qy 2348 GAGAGTTTTCAGAACATTTATTCAGCAAGTTTTCACAGATGTAATGAAGGCCAGGTAC 2407  
 Db 1670 AACCTGCGAGATGGGCTCCCTGGAGGTCCAGGAATGAGGGGCAATGC---CCGGAATC 1726  
 Qy 2408 CAGTCTTACTTCAGAGTGAAGAAATAGAAATTTGTATCATTTGCTCTCCCAACATGGCT 2467  
 Db 1727 CAGGAGNCCAGGAGTGTGGGAAACAGGCGCTCCCGAAGTCAAGGAGAAAGTGTCT 1786  
 Qy 2468 CCCGGGTATTCTGGGCCACTCTGGTCCGATAGGCCAGAGGGTCCAGAGGATTAACCTG 2527  
 Db 1787 GACCAGGTCTCTGGGCCATCTGGTCCCCGAGGTCAAGCTGTGTGTATGGCTTCCCGC 1846  
 Qy 2528 GTTTGCCAGGAGAGATGGTGTCTCTGATTTAGTGGGTGTCCCTGGAGTCCAGGTGCA 2587  
 Db 1847 GTCTAAAGGAAATGATGGTGTCTCTGTGTAAAGATGGAGAACGAGGTGGCCCTGGAGAC 1906  
 Qy 2588 GAGGATTAAGAGGCTACAGGAGAAATGGGGAAGAGGAGCAAGGCTTTGGGTATC 2647

Db 1907 CTGCGCTTCAGGCTCTCTCTGAAAGAAATGTGAACTGGACCTCAAGGACCCCAAGGC 1966  
 Qy 2648 CTGGAGAACAGGTCTCTCTGTCCTCCAGGTCCAGAGGGCCCTCTCTGGAATAGCAAG 2707  
 Db 1967 TACTTGGGCTTGG---TGGTCAAAAGGAGACACAGGACCCCTCTGGTCCCAAGGATTAC 2023  
 Qy 2708 AAGTCTCTTCCAGGAGACCCAGGTCTCTCTGCAAGATGAGACCATCGAAACCTGGAA 2767  
 Db 2024 AAGCTTCTGCTGTACAGGTGTCTCTCCAGGAGAAATGGAACCTGGGGAACAGGTC 2083  
 Qy 2768 TCCAAGGCAACAGGCCCCCAGG 2792  
 Db 2084 CAAAGGTCATCCCGGTGCACCTGG 2108

6

RESULT 10  
 US-10-177-293-67  
 ; Sequence 67, Application US/10:77293  
 ; Publication No. US2003012128A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Glatt, James  
 ; APPLICANT: Lillie, Karen  
 ; APPLICANT: Zhao, Xumei  
 ; APPLICANT: Ganmavaru, Manjula  
 ; APPLICANT: Kamatkar, Shubhangi  
 ; APPLICANT: Mertens, Maureen  
 ; APPLICANT: Myer, Vic  
 ; APPLICANT: Wang, Youzhen  
 ; APPLICANT: Xu, Yongyao  
 ; APPLICANT: Hoersch, Sebastian  
 ; APPLICANT: Monahan, John  
 ; APPLICANT: Meyers, Rachel E.  
 ; APPLICANT: Bast Jr., Robert C.  
 ; APPLICANT: Hortobagyi, Gabriel N.  
 ; APPLICANT: Pusztai, Lajos  
 ; APPLICANT: Meric, Funda  
 ; APPLICANT: Sahin, Aysegul  
 ; APPLICANT: Mills, Gordon B.  
 ; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,  
 ; OF INVENTION: PREVENTION, AND THERAPY OF BREAST CANCER  
 ; FILE REFERENCE: MR-038  
 ; CURRENT APPLICATION NUMBER: US/10/177,293  
 ; CURRENT FILING DATE: 2002-06-21  
 ; PRIOR APPLICATION NUMBER: US 60/299,887  
 ; PRIOR FILING DATE: 2001-06-21  
 ; PRIOR APPLICATION NUMBER: US 60/301,572  
 ; PRIOR FILING DATE: 2001-06-27  
 ; PRIOR APPLICATION NUMBER: US 60/306,501  
 ; PRIOR FILING DATE: 2001-07-18  
 ; PRIOR APPLICATION NUMBER: US 60/325,002  
 ; PRIOR FILING DATE: 2001-09-25  
 ; PRIOR APPLICATION NUMBER: US 60/362,585  
 ; PRIOR FILING DATE: 2002-03-05  
 ; PRIOR APPLICATION NUMBER: US 60/xxx,xxx  
 ; PRIOR FILING DATE: 2002-05-14  
 ; NUMBER OF SEQ ID NOS: 506  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 67  
 ; LENGTH: 5460  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; US-10-177-293-67

Query Match 8.5%; Score 242.6; DB 14; Length 5460;  
 Best Local Similarity 48.8%; Pred. No. 8.3e-56;  
 Matches 715; Conservative 0; Mismatches 744; Indels 6; Gaps 2;  
 Qy 1328 GTCTCCCGGAAACAGGACTTCAAGGCCCAAGGTGACCTGGACTGCTGGGAAC 1387  
 Db 650 GTCCCCCTGGTACATCTGTCTCTGTTTCTCTGATCTCCAGGATACCAAGGACCCC 709  
 Qy 1388 GTGGCTACCTGGACAACTGTGTCAAGATGTTAAGCTTGGATATCAGGGAATTCAGGGA 1447

1446	GACACCAGGTGTTCCAGGATCTCCAGGAATACAAGGAGCTCGAGGACTACACAGTTTACAA	1501
3513	CGCCCCGGACCTCAGGGTCTCTCGTGCAGCTGGAGAGAAAGGTGCTCTTGGAGAAA	3572
1506	AGGAGNACAGGGCGAGATGCTGACNAGGGTGATCTGGACTTCCTGGTTTTCCTGGGCT	1565
3573	AGGTCCCACAGGGGCTCGAGGAGAGATGGAGTTTCAAGGTCCTGTGGTCTCCACGGGCG	3632
1566	TATGGCATGCGAGATCAAAAGGTGAATGGGTGCCAAAGAGAGACAAAGGATCACTCG	1625
3633	AGCTGGTCTCGGGCTCCCTGGGAGACGGAGACNAGGTAATTGGTCAGCGGG	3692
1626	ATTTTATGGCAAAAAGGGTGCAAAAGGTGAAAGGGGAAATGCTGGCTTCCTGGCTCCC	1685
3693	ACAAAAGGACAGAGGGTGACAGGGAGAAATGGCCCTCCCGTCCCCCAGCTTCA	3752
1686	TGGACCTCTGSGANAACAGGAAGACATGGGAAGGAT-----GGATT	1727
3753	AGGACAGTTGGTCCCTGGAAATGCTGGAGGTGATGGTGAACACAGGTCCTAGAGACA	3812
1728	AATGGGTAGTCCGGTTTCAAGGAGACAGGATCCCTGGTGTCTCGGGGAGGATGG	1787
3813	GCAGGGATGTTTGGCAAAAGGTGATAGGTCGCCAGAGGCTTCTCTGGACCTCTCG	3872
1788	AACAGGGAGAGCTTGAATCCCAAGATTTCTTGGAAACCGAGGATTAATGGGCCAAA	1847
3873	TCCATAGTCTTTCAGGGTCTGCCAGGCCACCTGGTGCAAAAGGTGAAAATGGGATGT	3932
1848	GGGAGAAATGGGCTCCAGACAGCAAGGAAAAAAGAGGCCCCAGGATGCTTGTTT	1907
3933	TGGTCCATGATGGGCCACCTGGTCTTCAGGCCCAAGAGGCCCTCAAGTCCCAATGGAGC	3992
1908	AAATGGGAACAAATGGCTCACAGGCCAGCTCGAACACCGGGATCTAAGGGAAGCAAGG	1967
3993	TGATGGACCAAGGACCCCAAGTTCTGTGGTTTCAAGTTGGTGTGTGGAGAAAGGG	4052
1968	TGAACCTGGAATTAAGGGATGCTTGGGCTTCTGGGCTCAAGGGAGNACGAGGACAC	2027
4053	TGAACCTGGAGAACGAGGAACCCAGGACCTCTTGGGGNAGCAGGTGTAGCGGTGCCAA	4112
2028	GGTTCCTCCAGGAAACAGGATACATGGGTTTACCCTGGGATCAAGGAAAAAAGGGGA	2087
4113	AGGAGAAAGAGAGAGAAAGGGAGCTGTCTCACTGGAGCTCTGACCTCCAGGTGC	4172
2088	CAAGGAAATCAAGGTGAAAAGGTATTCAGGGTCAAAAAGGAGAAATGGAAGACAGG	2147
4173	CAAGGGGCCACAGGTGATGATGGGCCCTTAAGGGTAAACCGGGTCTCTGTGGTTTCTCG	4232
2148	AAATTCAGGGCAACAGGAAATTCAGGACCATCATGGTGCAAAAGGAGAGAGAGTGA	2207
4233	AGATCTGTCTCTTGGGAACTTGGCCCTGCAAGGTCAAGTGGTGTGGTGGTGACAA	4292
2208	GGGAGAACTGTGTCTCGAGGTGCCATTTGATCAAAAGGAGAAATCTGGGGTGGATG	2267
4293	GGGTCAAGATGAGATCCTGTGTAACCGGGTCTCTCGGCCCATCTGTGAGGCTGGCCC	4352
2268	GATGGGGCCGAGTCTTAAGGGCAACCTGGGATCCAGGTCTCTCAGGACCCCGAG	2327
4353	ACCAGGTCTCTCTGGAAAACGAGGTCTCTCTGGAGCTGCAAGGTGCAGAGGGAAGA	4412
2328	TTTGGATGGGAAGCCCGAAGAGAGTTTTTCAGAACAAATTTATTCGACAAAGTTTGCACAG	2386
4413	TGAAAAGGTGTAAGGGGAGCAGGTGCAGAGGTCTCTCTGGAAAACCGGCCCGT	4472
2387	ATGTAATAAGGCCAGCTACAGTCTTATTTCAAGTGGAGAAATTAGAAATTGTGATC	2446
4473	CGGTCTCTCAGGACCTGCAGGAAAGCCCTGGTCCAG-----AGGTCTTCGGGGC	4521
2447	ATTGCTGTCCCAACATGGCTCCCCGGGTATTCTCTGGGCCACCTGGTCTCGNATAGGCCAG	2506
4522	ATTCCTTGTTCC-TGTGGAGAACAGGTCTCTCTGGAGCTGCAGGCCAAGATGAGCAC	4579

Qy	2507	AGGGTCCACAGAGGATTACCTGGTTTCCACAGGAAGAGATGGTGTTCCTGGATTAAGTGGGTC	2556
Db	4580	CTGGTCTCTATGGCAACCTCTCTGGCTTACCTGGTCTCAAGAGTGACCTGGCTCCCAAGGGGTG	4639
Qy	2567	TCCTGGACGTCCAGCTGTCCAGAGGATTAAAGAGCCCTACCAAGGAAGAAATGGGGAAAAAAG	2626
Db	4640	AAAAGGACATCTCTGGTTTAATTGGCTGATTTGGTCTCCAGGAGAACAGGGGAAAAAG	4699
Qy	2627	GGAGCCAAAGGGTTTGGGTATCTCTGSAGAACAAAGTCTCTCTGGTCCCCCAGGTCCACAGG	2686
Db	4700	GTACCCAGAGGGCTC-----CCTGGAACTCAAGGATCTCCAGGAGCAAAAAGGGGATGGGG	4753
Qy	2687	GCCTCTCTGSANATCAAGAAAGAGGTCTCTCCAGGACCCAGTCTCCCTGGCAAAAGATG	2746
Db	4754	GAATTCCTGGTCTCT-----GCTGGTCCCTTAGTTCACGCTGTCTCCAGGCTTACCAG	4807
Qy	2747	GAGACCATGAAACCTGGAATCCAAGGGCAACGAGGCCGCCCCACAGG	2793
Db	4808	GTCTGAGAGCCCAAGGGTTACAAAGGCTCTACTGGACCCGTGSC	4854
RESULT 9			
US-10-301-822-32			
; Sequence 32, Application US/10301822			
; Publication: NO. US20030148410A1			
; GENERAL INFORMATION:			
; APPLICANT: Millennium Pharmaceuticals, Inc.			
; APPLICANT: Berger, Allison			
; APPLICANT: Guillemette, Tracy L.			
; APPLICANT: Kamatkar, Shubhangi			
; APPLICANT: Schlegel, Robert			
; APPLICANT: Monahan, John E.			
; APPLICANT: Thibodeau, Stephen N.			
; APPLICANT: BURGART, Lawrence J.			
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND			
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND			
; TITLE OF INVENTION: THERAPY OF COLON CANCER			
; FILE REFERENCE: MP001-0292PRNM			
; CURRENT APPLICATION NUMBER: US/0301,822			
; CURRENT FILING DATE: 2002-11-21			
; PRIOR APPLICATION NUMBER: US 60/339,971			
; PRIOR FILING DATE: 2001-12-10			
; PRIOR APPLICATION NUMBER: US 60/361,978			
; PRIOR FILING DATE: 2002-03-05			
; PRIOR APPLICATION NUMBER: US 60/361,988			
; PRIOR FILING DATE: 2002-05-20			
; NUMBER OF SEQ ID NOS: 228			
; SOFTWARE: FastSeq for Windows Version 4.0			
; SEQ ID NO 32			
; LENGTH: 5460			
; TYPE: DNA			
; ORGANISM: Homo Sapiens			
; FEATURE:			
; NAME/KEY: CDS			
; LOCATION: (103)...(4503)			
US-10-301-822-32			
Query Match			
Best Local Similarity 8.5%; Score 242.6; DB 12; Length 5460;			
Matches 715; Conservative 0; Mismatches 744; Indels 6; Gaps 2;			
Qy	1328	GTCTCCGGGAAAACACGAGGACTTCAAGGCCCAAGAGTGACCTGGACTGCTTGGGAACC	1387
Db	650	GTCCCCCTGTGTACATCTGGTGATCTCTGGTTCCCTCGATCTCCAGNATCAAGGACCCC	709
Qy	1388	CTGGCTACCTCGGACAACTGGTCAAGATGGTAAGCCTGGATATCAGGGGAATTCACAGGA	1447
Db	710	CTGTGTGAACCTCGGGCAAGCTGGTCTCTCAGGCCCTCCAGGACCTCTCTGGTGTCTATAGTGC	769
Qy	1448	CACAGGTGTCTCAGAGATCTCCAGGAAATCAAGAGCTCAGGAGCTACCAAGGTTCACAAAG	1507
Db	770	CATCTGTCTCTGTGGAAAAAGATGGGAATCAGGTAGACCCGGACGACCTGGAGAGCGAG	829



:	TYPE: DNA					
:	ORGANISM: Homo sapiens					
:	US-10-198-846-13890					
<hr/>						
	Query Match	8.5%	Score 244.6;	DB 14;	Length 6545;	
	Best Local Similarity	51.0%;	Pred. No. 2.6e-56;			
	Matches 759;	Conservative 0;	Mismatches 684;	Indels 44;	Gaps 6;	
<hr/>						
QY	1326	TTCCTCCTCCGGGAAACACAGGACTTCAAGGCCCCCAAAGGTGAOCCTCGACTGCCTGSGAA	1385			
Db	3432	TGGAGCTCAGGGTGCACTGCATGAAAGGAGGGGAAGGTCCCAGGCCCAACGAGTCC	3491			
QY	1386	CCCTGGCTACCCTGGACAACCTGCTCAAGATGGTAAGCCTGGATATCAGGGAATTGCAGG	1445			
Db	3492	AGTTGGCTCACAGGAGAAGCTGGGTACAGAGTACAGCTGGCCCCAATTGTTTTACCAGG	3551			
QY	1446	GACACCAAGGTGTTCTCAGGATCTCCAGGAATACAAAGGAGCTCGAGGACTACAGGTTACAA	1505			
Db	3552	CGCCCCGGACCTCAGGCTCCTCTGGTCCAGCTGGAGAGAAAGTCTCTCGAGAAAA	3611			
QY	1506	AGGAGAACCAAGGCGAGATGGTGACAGGGTGATCTGTGGAATTCCTGGTTTTCTGGGCT	1565			
Db	3612	AGGTCCCAAGGGCTCGAGGAGAGATGGATTCAAGGTCCTGTTGGTCTCCCAAGGCC	3671			
QY	1566	TCATGGCATGTCAGGATCAAAAGGCTGAAATGGGTGCCAAAAGGACAAAAGGATCACCTGG	1625			
Db	3672	AGCTGGTCTCCCGCTCCCCTGGGGAAGACGGAGACAAGGTGAAATGGTAGCCCGG	3731			
QY	1626	ATTTTATGGGNAAGAGGTGCAAAAGGTGAAAAGGGGAATGCTGGCTTCCCTGGCCTCCC	1685			
Db	3732	ACAAAAGGCAGCAAGGTTGCAAGGAGAGAAATGGCCCTCCCGGTCCCCCAAGTCTTCA	3791			
QY	1686	TGCACCTGCTGGAGAACACAGGAACACATGAAAAGGAT-----GGATT	1727			
Db	3792	AGGACAGATTGGTCCCTCGAATTGCTGGAGGTGATGGTGAACACAGTCTACAGGACA	3851			
QY	1728	AATGGGTAGTCCCGTTTCAAGGAGAGCAAGGATCCCTGGTGTCTCCGGGGCAGATGG	1787			
Db	3852	GCAGGGATGTTTGGGCAAAAGGTGATAGGGTGCCAGAGGCTTCCTCGACCTCCTCG	3911			
QY	1788	AACACGGGGAGAGCTGGATCCCCAGGATTTCTTGGAAACCGAGGATTAATGGSCCAAA	1847			
Db	3912	TCCAATAGTCTTCAGGGTCTGCCAGGCCCCACCTGGTGAAAAGGTGAAATGGGATGT	3971			
QY	1848	GGGAGAAATTGGCTCTCAGACACAGCAAGSAAAAAGAGACCCACAGGATGCCTGGTTT	1907			
Db	3972	TGGTCCCATGGGCCACCTGGTCTCCAGGCCCAAGAGGCCCTCAAGGTCCCAATGAGC	4031			
QY	1908	AATGGGAAGCAATGGCTCACAGGCGCAGCTTGAAACACCGGGATCTAAGGGAAGCAAGG	1967			
Db	4032	TGATGGACCAACAGGAGCCCCAGGGTCTGTTGGTTCAAGTTGGTGGTGTGGAGAAAAAGG	4091			
QY	1968	TGAACCTGGAAATTCAGGGGATGCCCTGGGGCTTCTGGGCTCAAGGGAGAACACAGGAGCAAC	2027			
Db	4092	TGAACTGGGAGAACAGGGAACCCAGGSCCTCTTGGGAGCAGGTTGAGCGGTCCCAA	4151			
QY	2028	GGTTTCCGAGGAGAACAGAGTACATGGGTTTACCGGGGANTTCAAGAAAAAGGGGGA	2087			
Db	4152	AGGAGAAGAGGAGAGAAAGGGGAAGCTGTGTCACCTGGAGCTCTCGAACCTCCAGGTGC	4211			
QY	2088	CMAAGGAAATCAAGTGAAAAAGGTATTCAGGGTCAAAAGGGAGAAATGGAACACAGGG	2147			
Db	4212	CAAGGGCCACAGGTGATGATGGCCCTAAGGTTAA CCGGGTCTGTTGGTTTCTCG	4271			
QY	2148	AATTCCAGGGCAACAGGGAATTCAGGCCCATCATGGTGCAAAAGGAGAGAGAGTGAATA	2207			
Db	4272	AGATCTCTGGTCTCTCTGGGGAACCTGGCCCTGCAGGTCAAGTGGTGTGGTGTGACAA	4331			
QY	2208	GGGAGAACCTGGTCTCGAGGTGCCATTGGATCAAAAGGAGAAATCTGGGTGGATGGCTT	2267			
Db	4332	GGGTGAAGATGGAGATCTGTGTCAACCGGGTCTCTCTGGCCCCATCTGTGAGGCTGGCCC	4391			
QY	2268	GATGGGGCCCGCAGGTCTTAAGGGGGCAACCTGGGGATCCAGGTCTCAGGAGACCCCGAGG	2327			

```
Db 4156 GGAATCCCGAGGAGAAAGGCAAGAGGGCCAGAGATGGAAAGCCGGGTCCCCCTGGAGAG 4215
Qy 1747 AAGGAGAGAGAGGATCCCTCGTGTCTCCGGGCGAGGATGGAAACCGGGAGAGCTCGA 1806
Db 4216 CCGGCAAGAGCAGGAGAGCAGGTCTACAGAGCAGAGGGTCCCGGAGGCCACTGGC 4275
Qy 1807 ATCCAGGATTTCTTGAAACCGAGGATTAATGGGCCAAAGGGAGAAATTGGGCTCA 1866
Db 4276 TTAAGGACACACAGGCGATTCTGGTGACCCGGTCCCGGGAGAGTCTGGTCAATG 4335
Qy 1867 GCACAGCAAGGAAAGAGAGCCCGAGGATCCCTGGTTTAATGGAGAGCAATGGCTCA 1926
Db 4336 GGGCTTCTCGTCAGGAAGGGTTACAGGAAAGATGGTGACACTGGACCCACTGGCCA 4395
Qy 1927 CAGAGCCAGCTGGAACACCGGATCTAAGGGAAGCAAAAGGTCAACTCGAATTCAAGG 1986
Db 4396 CAGGTCCCAAGGACCAAGGGCCACCGGCNAGATGGATCACCGGATCTCCAGA 4455
Qy 1987 ATGCTGGGCTTCTGGCTCAAGGGAGAACCGAGGACCGGTTCCCGAGAGAACCA 2046
Db 4456 GAGCTTCCGCTTCCGCTGCGGTCCTCGGCGCTCCCGGAGAACCGAGGAGAAAGAGTC 4575
Qy 2047 GATCATATGGTTTACCGGGATTCAAGGAAAGAGGGGCAAAAGCAATCAAGTGAA 2106
Db 4516 GGAATTCCTGGCTTCTGGTCCCTCGGCTCCCGGAGAACCGAGGAGAAAGAGTC 4575
Qy 2107 AAGGTATTCAAGGTCAAAGGAGAGAAATGGAGACAGGGAATCCAGGGCAACAGGGA 2166
Db 4576 CAGGCAAGAGGGGGTCTCTGGAGACCTGGAGAGCTGGATTCAAAGGAGAAAGGGA 4635
Qy 2167 ATTCAAGGGCCATCATGTTGCAAAAGGAGAGAGGTGAAAGGGAGAACTGGTGTCCA 2226
Db 4636 GATCTCTGGATCAAGAGTGACAAAGGACCTCTCGTGGAAAGGCCAGCTGGGACCT 4695
Qy 2227 GGTGCCATTGGATCAAGAGAGATCTGGGGTGGATGGCTTGATGGGGCCCGAGTCT 2286
Db 4696 GGAATCCAGGCCACAAAGGCCACAGAGGCTGATGGGTCCCAAGGACTACTCTGGGAG 4755
Qy 2287 AAGGGCAACTCTGGGATCCAGTCTCTCAGGACCCCGAGGTTTGGATGGAGCCCGGA 2346
Db 4756 AATGGACAGTTGGACCCCGAGGCTCCAGGCCAGCGGGATTTCCAGGACTGAGGGG 4815
Qy 2347 AGAGATTTTCAGAAACAATTTATTCAGAAAGTTTGCACAGATGTAATAAGAGCCAGCTA 2406
Db 4816 GAGTCTCCATCCATGGAAGCCCTCGCTCGGCTT-----ATTCAAGAGAGAGCTGGGAA 4868
Qy 2407 CCAGTCTTACTTCAGAGTGAAGAAATTAGAAATTGTGATCATTTGGCTGTCCCAACATGC 2466
Db 4869 GCAGCTTGAAACAGACTCGCTACCTCTCTGGCCAGATGCCCGCGGTACATGAGTC 4929
Qy 2467 -TCCCCGGGTATTCTCTGGGCCAGCTGTGTCGATAGGCCAGAGGGTCCAGAGGATTACC 2525
Db 4929 ATCTCAAGGACAGACTGGGCCCCCGAGGGCCCCCTGGAAAGATGGGCTTCCAGGCCGGGC 4988
Qy 2526 TGGTTTCCAGGAGAGATGGTGTCTGTGATAGTGGGTGTCCTCTGGAGCTCAGGTGT 2585
Db 4989 CGGCCCCATCGGGGAGCCAGGTCTGTGGAGGGGGGTCTGGAAAGGACCTCTCGACC 5048
Qy 2586 CAGAGGATTAAGAGGCTTACAGGAGAAATGGGAAAGAGGGAGCCAGGGTTTCGGTA 2645
Db 5049 CATAGTTCCAAGGTGAGGAGGAGCCAAAGGTGACCCAGGTGACCTGGAGTTGGCCT 5108
Qy 2646 TCCTGGAGAAACAAGGTCTCTGTGTCCTCGAGTCCAGAGGGCCCTCTGGAATAAGCAA 2705
Db 5109 CCAGGGAGATGGACCCCTGGAATCCAGGTCAACCCGGGNACTTGGCTATCTAA 5169
Qy 2706 AGAAGTCTCCAGGAGAGCCAGGTCTCTCTGGCAAGATGGAGACCATGGAAACCTGG 2765
Db 5169 AGATGGACTTCTGGGATCTCTGGCTCCCTCAAGGGGAGACAGGACCACTGGACATCTCG 5228
Qy 2766 AATCCAGGGCCACAGGCCCCCGAGGATCTCGACCCATCACTATG 2814
Db 5229 CCTCCAGGACCTCCCGTCCCCCGAGGCCAATGTGACCCCTTCCAGTGT 5277
```

## RESULT 6

```
US-09-918-995-17490
; Sequence 17490, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO: 17490
; LENGTH: 496
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(496)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-17490
```

Query Match 9.3%; Score 267; DB 11; Length 496;

Best Local Similarity 100.0%; Pred. No. 2.3e-63;

Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy 2599 GGCCTACCCAGGAAGAAATGGGAAAGAGGGAGCCAAAGGTTTGGGTATCTCGGAGAACAA 2658
Db 85 GGCCTACCCAGGAAGAAATGGGAAAGAGGGAGCCAAAGGTTTGGGTATCTCGGAGAACAA 144
Qy 2659 GGTCTCTCTGTGTCCTCCAGGTCCAGAGGGCCCTCTGGAATAAGCAAGAGGTCTCTCCA 2718
Db 145 GGTCTCTCTGTGTCCTCCAGGTCCAGAGGGCCCTCTGGAATAAGCAAGAGGTCTCTCCA 204
Qy 2719 GGAGACCCAGGTCTCCCTGGCAAAAGATGGAGACCATGGAAACCTGGAATCCAAAGGGCAA 2779
Db 205 GGAGACCCAGGTCTCCCTGGCAAAAGATGGAGACCATGGAAACCTGGAATCCAAAGGGCAA 264
Qy 2779 CGAGCCCCCAGGAGCATCTCGACCCATCACTATGTTTTAGTGTAAATTCGCAGAGAGAT 2838
Db 265 CGAGGGCCCCCAGGAGCATCTCGACCCATCACTATGTTTTAGTGTAAATTCGCAGAGAGAT 324
Qy 2839 CCGTTTCAGAAAGGACCAAACTATTAG 2865
Db 325 CCGTTTCAGAAAGGACCAAACTATTAG 351
```

## RESULT 7

```
US-10-198-846-13890
; Sequence 13890, Application US/10198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
; APPLICANT: Lilly, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; THERAPY OF BREAST CANCER
; FILE REFERENCE: YRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14094
; SOFTWARE: FastSeq for Windows Version: 4.0
; SEQ ID NO: 13890
; LENGTH: 6545
```

Db	3529	GGTGAGGTTGGCGAGAAAGGTGATCAGGGTTCATCTGGAGTTCAGGTGTTTCATGTGGGCCC	3538
Qy	1567	CATGGCATGCCAGCATCAAAAGGGTGAAATGGTCCCAAAGGAGACAAAGGATCACCTGGGA	1626
Db	3589	CCAGGAAACCCGGGCACACAGGGGCAGATGAATTCAGAGAGCTGCTGACACACAGGA	3648
Qy	1627	TTTTATGGCAAAAAGGTGCAAAAGGTGAAAGGGGAATGCTGCTTCCCTGGGCTCCCT	1686
Db	3649	ATCCAAGGTCACCTGGMAAAGGCCCTCCTGGCCCCCAAGCCCATCTGGATTACCC	3708
Qy	1687	GGACCTCTGGAGAACACAGGACATGGAAGATGGATTAAATGGGTAGTATGCCGTTTC	1746
Db	3709	GGAAATCCAGGAGAAGAGGCAAAAGAGGGCAGAGATGGAAGCCGGTCCCTCTGGAGAG	3768
Qy	1747	AAGGGAGAGCAGGATCCCTGGTGTCTCCGGGCGAGGTGGAACACGGGAGAGCCTGGGA	1806
Db	3769	CCGGCAAGCAGAGAGCCAGTCTACAGGACACAGGGTGCAGAGGCCACACTGGC	3828
Qy	1807	ATCCAGGATTTCTCTGMAAACCGAGGATTAATGGGCCAAAAGGAGAAATTTGGCCCTCCA	1866
Db	3829	TTCAAGGGAACACACAGGCGATTCTGGTGCAACCCGCTCCCGGGGAGAGTCTGGTGCCATG	3888
Qy	1867	GGACAGCAAGAAAAAAGGAGCCCCCAGGGATGCTGTTTAATGGGAAGCAATGGCTCA	1926
Db	3889	GGGCTTCTGGTCAGGAAGGGTTACCAGGAAGAATGGTGACATGGACCCACTGGGCCA	3948
Qy	1927	CCAGGCAGCCTCGAAACCCGGGATCTAAAGGAAGCAAGGTGAACCTGGAATTTCAAGGS	1986
Db	3949	CAGGGTCCCCAAGAACACAGGGGCCACCGGGCAAGATGGATCACCGGATCTCCAGGA	4008
Qy	1987	ATGCTCTGGGCTTCTGGCTCAAGGAGAAACAGGAGCAACGGGTTTCCCGAGGAAACCA	2046
Db	4009	GAGCCTGGCCCTTCAGGAACCCCTGGCCAGAAAGCAAGCAAAAGGGGAAATGGCAGCCCA	4068
Qy	2047	GGATACATGGTTTACCCGGGATTTAAGGAAAAAAGGGGCAAAAGGAAATCAAGGTGAA	2106
Db	4069	GGACTTCTGGCTTCTGGGTCCCGCTGGGCTCCCGGAGAACCCAGGAGAAAGGAGTGC	4128
Qy	2107	AAAGGTATTGAGGTCAAAAGGGAGAAAAATGSAAGACAGGGAATTCAGAGGCCAACAGGA	2166
Db	4129	CCAGGCAAGAGGGGGTCCCTGGGAAGCCTGGAGAGCTGGATTCAANGAGAAAGGGGA	4188
Qy	2167	ATCCAAGGCCATCATGTGTCAAAAGGAGAGAGGTGAAAGGGGAGAAACCTGTGTCCGA	2226
Db	4189	GATCCTGGGATCAAAAGGTGACAAAGGACCTCCTGTGTAAGAGCCAGCCTGGGACCT	4248
Qy	2227	GGTGCCATTGGATCAAAAGGAGATCTGGGGTGGATGGCTTGATGGGGCCGCGAGTCT	2286
Db	4249	GSAATCCAGGCCCAAGAGGCCACACAGGCCCTGATGGTCCCCAAGGACTACCTGGGAG	4308
Qy	2287	AAGGGCAACCTCGGGATCCAGGTCTTCAGGAGCCCCCAGGTTTGGATGGGAAGCCCGGA	2346
Db	4309	AATGGAACGATTGGAACCCCAAGGCCCTTCAGGCCAGCGGGATTTCCAGAGCTAGAGGGG	4368
Qy	2347	AGAGAGTTTTCAGAACAAATTTATTCGAAGTTTGCACAGATGTAATAAGAGCCCACTA	2406
Db	4369	GAGTCTCCATCCATGGAACCCCTCGCTGGCTT-----ATTCAAGAGAGCTGGGAA	4421
Qy	2407	CCAGTCTTACTTCAGATGGAGAAATTAGAAATTTGTGATCATTCCTGTGCCAACATGGC	2466
Db	4422	CGACCTTGAACCCAGACTCGCTTACCTCTGGCCAGATGCCCGCGGTACATGAAGTC	4481
Qy	2467	-TCCCGGGGATTCTCTGGGCCACCTGGTCCGATAGGCCCCAGAGGGTCCCAGAGGATTACC	2525
Db	4482	ATCTCAAGGAGACTGGGCCCCCAAGGCCCTTCGAAAGAGATGGGCTTCCAGGCCGGGC	4541
Qy	2526	TGGTTTGCAGGAAGAGATGGTGTTCCTGGATTAGTGGGTCTCCCTGGACGTCCAGGTGT	2585
Db	4542	CGGCCCATGGGGAGCCAGGTGCTGCTGGGACAGGGGGTCTGGAAGGACCTCTGGACC	4601
Qy	2586	CAGAGGATTTAAAGGCCCTTACAGGAAGAAATGGGGGAAAGAGGAGCCAGGGTTTGGGT	2645

```

Db      4602  CATAGGTCCAAAGGTGAGCGAGAGCCAAAGGTGACCCAGGTGCACCTGGAGTTGGCCT  4661
Qy      2646  TCTTGGAGAAACAAGGTCCTCTCGTCCCGCCAGGTCCAGAGAGGGCCCTCTCGAATAAGCAA  2705
Db      4662  CCGAGCGAGATGGGACCCCTCGAATCCAGGTCAACCCGGGGAACTGGCTATGCTAA  4721
Qy      2706  AGAAGGTCTCTCAGGAGACCCAGGTCTCCTGCGCAAGATGGAGACCATGGAAAACCTGG  2765
Db      4722  AGATGGACTTCTTGGATCCCTGGCCCTCAAGGGGAGACAGGACCCAGCTGGACATCCTGG  4781
Qy      2766  AATCCCAAGGCAACCCAGGCCCCCGCCAGGCGATCTCGACCCCATCACTATGT  2814
Db      47829 CCTCCCAAGACCTCCCGGTCCCTCCCAAGGCGCAATGTGACCCCTTCCCACTGT  4830

RESULT 5
US-10-202-167-1
; Sequence i, Application US/10202167
; Publication No. US20030143564A1
; GENERAL INFORMATION:
; APPLICANT: Burgeson, Robert Eugene
; APPLICANT: Koch, Manuel
; APPLICANT: Bruckner-Ruderman, Leena
; APPLICANT: Keene, Douglas R.
; APPLICANT: Brunken, William Joseph
; TITLE OF INVENTION: COLLAGEN XXII, A NOVEL HUMAN COLLAGEN
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: 10287-072061
; CURRENT APPLICATION NUMBER: US/10/202,167
; PRIOR FILING DATE: 2003-03-31
; PRIOR APPLICATION NUMBER: US 60/309,158
; PRIOR FILING DATE: 2001-07-31
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 6352
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (448) ... (5325)
; US-10-202-167-1

```

•

QY 2224 CGAGTGCCTTGGATCAAAAGGAGAAATCTGGGTGGATGGCTTGTAGTGGGCGCCGCGAGGT 2283  
DB 511 CGAGTGCCTTGGATCAAAAGGAGAAATCTGGGTGGATGGCTTGTAGTGGGCGCCGCGAGGT 570  
QY 2284 CTTAAGGGGCAACCTGGGGATCCAGGTCTCAGGAGCCCCCAGGTTTGGATGGGAAGCCC 2343  
DB 571 CTTAAGGGGCAACCTGGGGATCCAGGTCTCAGGAGCCCCCAGGTTTGGATGGGAAGCCC 630  
QY 2344 GGAAGAGAGTTTTCAGAACAAATTTATTCCGCAAGTTTGCACAGATGTAAATAAGAGCCAG 2403  
DB 631 GGAAGAGAGTTTTCAGAACAAATTTATTCCGCAAGTTTGCACAGATGTAAATAAGAGCCAG 690  
QY 2404 CTACAGCTTTACTTTCAGAGTGGAGAAATTAGAAATTTGATCATTCCTGTGCCAACAT 2463  
DB 691 CTACAGCTTTACTTTCAGAGTGGAGAAATTAGAAATTTGATCATTCCTGTGCCAACAT 750  
QY 2464 GGCTCCCGGGTATTCCTGGGGCCACCTGGTCCGATAGCCCGAGGCTCCAGAGGATTA 2523  
DB 751 GGCTCCCGGGTATTCCTGGGGCCACCTGGTCCGATAGCCCGAGGCTCCAGAGGATTA 810  
QY 2524 CCTGGTTTCCAGGAAGAGATGGTTCCTCGATTAGTGGGTGTCCTCGACGTCCAGGT 2583  
DB 811 CCTGGTTTCCAGGAAGAGATGGTTCCTCGATTAGTGGGTGTCCTCGACGTCCAGGT 870  
QY 2584 GTCAGAGATTAAGGCTTACAGGAGAAATGGGAAAGAGGAGCCAAAGGTTTGGG 2643  
DB 871 GTCAGAGATTAAGGCTTACAGGAGAAATGGGAAAGAGGAGCCAAAGGTTTGGG 930  
QY 2644 TATCTCGAGACAGAGTCTCTCGTCCCGCCAGGTCAGAGGCGCTCCTGGGAATAGC 2703  
DB 931 TATCTCGAGACAGAGTCTCTCGTCCCGCCAGGTCAGAGGCGCTCCTGGGAATAGC 990  
QY 2704 AAAGAAGTCTCTCCAGGAGACCCAGGTCTCCTGGCAAGATGGAGACCATGGAAGAACCT 2763  
DB 991 AAAGAAGTCTCTCCAGGAGACCCAGGTCTCCTGGCAAGATGGAGACCATGGAAGAACCT 1050  
QY 2764 GGAATCCAAAGGCAACAGGCCCCCAGGATCTCGACCATCACTATGTTTAGTGTA 2823  
DB 1051 GGAATCCAAAGGCAACAGGCCCCCAGGATCTCGACCATCACTATGTTTAGTGTA 1110  
QY 2824 ATTGCAGAGAGATCCGTTTCAGAAAGGACCAAACTATTAG 2865  
DB 1111 ATTGCAGAGAGATCCGTTTCAGAAAGGACCAAACTATTAG 1152

## RESULT 3

US-09-918-995-33525  
; Sequence 33525, Application US/09918995  
; Publication No. US20030073623A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; FILE REFERENCE: 20411-756  
; CURRENT APPLICATION NUMBER: US/09/918,995  
; PRIOR FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US/09/235,076  
; PRIOR FILING DATE: 1999-01-20  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 33525  
; LENGTH: 412  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: (1)..(412)  
; OTHER INFORMATION: n = A,T,C or G

Query Match 13.3%; Score 381; DB 11; Length 412;  
Best Local Similarity 97.8%; Pred. No. 4.3e-95;

Matches 400; Conservative 0; Mismatches 0; Indels 9; Gaps 2;  
QY 1103 AACAAATTGAAAACAGCCCTTACATCCAGTTTTCAGGATCTTATCAATGGGAAACCC 1162  
DB 3 AACAAATTGAAAACAGCCCTTACATCCAGTTTTCAGGATCTTATCAATGGGAAACCC 62  
QY 1163 AAATTGGAAAATATTCTGGAAGAAAGAACTGTTCAGTTTGATGTCGAAAGTTGCGAA 1222  
DB 63 AAATTGGAAAATATTCTGGAAGAAAGAACTGTTCAGTTTGATGTCGAAAGTTGCGAA 122  
QY 1223 TCTACTGTGACCCAGAAACAGACACCGGAGACAGCATGTGAGATTCCTGATTTAATG 1278  
DB 123 TCTACTGTGACCCAGAAACAGAAACACCGGAGACAGCATGTGAGATTCCTGATTTAATG 182  
QY 1279 -----TGCTTAAATGTCCTAGTGTAGGTTCAACTCCAGCTCCCTGTGATTTGTCCTC 1333  
DB 183 GAGAGTGCTTAAATGTCCTAGTGTAGGTTCAACTCCAGCTCCCTGTGATTTGTCCTC 242  
QY 1334 CGGGAACACAGGACTTCAGGCCCCAAAGGTGACCTTGGACTGCTGGGAACCCCTGGCT 1393  
DB 243 CGGGAACACAGGACTTCAGGCCCCAAAGGTGACCTTGGACTGCTGGGAACCCCTGGCT 302  
QY 1394 ACCCTGGACAACTCTGTCAGATGTCAGCTTGGATATCAGGGAATTCAGGACACCCAG 1453  
DB 303 ACCCTGGACAACTCTGTCAGATGTCAGCTTGGATATCAGGGAATTCAGGACACCCAG 362  
QY 1454 GTGTTCCAGGATCTCCAGGAATACAGGAGCTTCGAGGACTACCAAGTTA 1502  
DB 363 GTGTTCCAGGATCTCCAGGAATACAGGAGCTTCGAGGACTACCAAGTTA 411

## RESULT 4

US-10-202-167-3  
; Sequence 3, Application US/10202167  
; Publication No. US20030143564A1  
; GENERAL INFORMATION:  
; APPLICANT: Burgess, Robert Eugene  
; APPLICANT: Koch, Manuel  
; APPLICANT: Bruckner-Tuderman, Leena  
; APPLICANT: Keene, Douglas R.  
; APPLICANT: Brunken, William Joseph  
; TITLE OF INVENTION: COLLAGEN XXII, A NOVEL HUMAN COLLAGEN  
; TITLE OF INVENTION: AND USES THEREOF  
; FILE REFERENCE: 10287-072001  
; CURRENT APPLICATION NUMBER: US/10/202,167  
; PRIOR FILING DATE: 2003-03-31  
; PRIOR APPLICATION NUMBER: US 60/309,158  
; PRIOR FILING DATE: 2001-07-31  
; NUMBER OF SEQ ID NOS: 3  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 4881  
; TYPE: DNA  
; ORGANISM: Homo sapiens

Query Match 9.9%; Score 285; DB 12; Length 4881;  
Best Local Similarity 50.4%; Pred. No. 1.3e-67;

Matches 751; Conservative 0; Mismatches 730; Indels 8; Gaps 2;  
QY 1327 TGTCTCCCGGAAAACCCAGGACTTCAAGGCCCCCAAGAGTGACCCCTGGACTGCTGGGAAC 1386  
DB 3349 TGGCTCTCTGGCCCCCAGGGCTTCCCTGCTTACCAGTTTAAAGGGGACAAAGGTGTC 3408  
QY 1387 CTGGCTACCTGGACAACTGGTCAAGATGTAGCCCTGGATATCAGGGGAATTCAGGG 1446  
DB 3409 CCAGGAAGCCAGGAGAGAGGACAGAGGAAAAAGGAGAGGCTGGGCTCCAGGC 3468  
QY 1447 ACACAGGTGTTCCAGGATCTCCAGGATACAGGAGCTCGAGGACTACCAAGTTACAA 1506  
DB 3469 CTACAGGSCCCCGAGGAATAGCTGGACCAAGGGAAGTCAAGGAGAACGTGTGCAGAT 3528  
QY 1506 CGAGAGCCAGGCGAGATGTTGACAAGGTTGATCTGCACTTCTCTGGGCTT 1566

```
QY 1824 AACCGAGGATTAATGGGCCCAAAAGGAGAAATTTGGGCTCCAGGACAGCAAGGAAAAA 1883
DB |||||||
QY 249 AAA CGAGGATTAATGGGCCCAAAAGGAGAAATTTGGGCTCCAGGACAGCAAGGAAAAA 308
DB |||||||
QY 1884 AGGAGCCCCAGGATGCTGCTGTTTAATGGGAAGCAATGGCTCACAGGCGAGCCTGGAAC 1943
DB |||||||
QY 309 AGGAGCCCCAGGATGCTGCTGTTTAATGGGAAGCAATGGCTCACAGGCGAGCCTGGAAC 368
DB |||||||
QY 1944 ACCGGGATCTAAGGGAAGCAAGGTGAACCTGGAAATTCAGGGGATGCTGGGGCTCTCG 2003
DB |||||||
QY 369 ACCGGGATCTAAGGGAAGCAAGGTGAACCTGGAAATTCAGGGGATGCTGGGGCTCTCG 428
DB |||||||
QY 2004 GCTCAAGGGAGAACAGGAGCAACGGGTTCGCCAGGAAACAGGATACATGGGTTTACC 2063
DB |||||||
QY 429 GCTCAAGGGAGAACAGGAGCAACGGGTTCGCCAGGAAACAGGATACATGGGTTTACC 489
DB |||||||
QY 2064 CGGATTCAGGAAAAAGGGGACAAAGGAATTCAGGTCAGGAAAAAGGATTCAGGGTCA 2123
DB |||||||
QY 489 CGGATTCAGGAAAAAGGGGACAAAGGAATTCAGGTCAGGAAAAAGGATTCAGGGTCA 548
DB |||||||
QY 2124 AAAGGGAGAAAAATGAAGACAGGGGAATTCAGGGGCAACAGGGAATTCAGGCCATCATGG 2183
DB |||||||
QY 549 AAAGGGAGAAAAATGAAGACAGGGGAATTCAGGGGCAACAGGGAATTCAGGCCATCATGG 608
DB |||||||
QY 2184 TGCAAAAGGAGAGAGAGGTGAAAGGGAGAACCTGGTTCGAGGTGCCAATGGATCAAA 2243
DB |||||||
QY 609 TGCAAAAGGAGAGAGAGGTGAAAGGGAGAACCTGGTTCGAGGTGCCAATGGATCAAA 668
DB |||||||
QY 2244 AGGAAATCTGGGTGGATGCTTCATGGGCGCGCAGGTCTTAAGGGGCAACCTGGGGA 2303
DB |||||||
QY 669 AGGAAATCTGGGTGGATGCTTCATGGGCGCGCAGGTCTTAAGGGGCAACCTGGGGA 729
DB |||||||
QY 2304 TCCAGGTCTCAGGACCCCCAGGTTTGGATGGGAAGCCCGGAGAGAGTTTTCAGAAACA 2363
DB |||||||
QY 729 TCCAGGTCTCAGGACCCCCAGGTTTGGATGGGAAGCCCGGAGAGAGTTTTCAGAAACA 789
DB |||||||
QY 2364 ATTATTCGACAAAGTTTCGACAGATGTAATAGAGCCCGACGCTACAGTCTTACTTTCAGAG 2423
DB |||||||
QY 789 ATTATTCGACAAAGTTTTCGACAGATGTAATAGAGCCCGACGCTACAGTCTTACTTTCAGAG 848
DB |||||||
QY 2424 TGGAGAAATTAAGAAATTTGATCATGCTGCTCCCAACATGCTCCCGGCTATTCTCTGG 2483
DB |||||||
QY 849 TGGAGAAATTAAGAAATTTGATCATGCTGCTCCCAACATGCTCCCGGCTATTCTCTGG 908
DB |||||||
QY 2484 GCCACTGCTCGATAGGCCAGAGGGTCCAGAGSATTACCTGGTTTGCAGGAAGAGA 2543
DB |||||||
QY 909 GCCACTGCTCGATAGGCCAGAGGGTCCAGAGSATTACCTGGTTTGCAGGAAGAGA 968
DB |||||||
QY 2544 TGGTGTCTCGATTAAGTGGGTGCTCCCTGGAGGTCAGGTGTCAGAGATTAAAAGGCT 2623
DB |||||||
QY 969 TGGTGTCTCGATTAAGTGGGTGCTCCCTGGAGGTCAGGTGTCAGAGATTAAAAGGCT 1028
DB |||||||
QY 2604 ACCAGGAAGAAATGGGGAAGGAGCAAGGGTTTGGGTATCTCGAGAACAGGTGCC 2663
DB |||||||
QY 1029 ACCAGGAAGAAATGGGGAAGGAGCAAGGGTTTGGGTATCTCGAGAACAGGTGCC 1088
DB |||||||
QY 2664 TCTGTGTCCTCCAGGTCAGAGGGCCCTCTCGGAATAAGCAAGAGTCTCTCAGGAGA 2723
DB |||||||
QY 1089 TCTGTGTCCTCCAGGTCAGAGGGCCCTCTCGGAATAAGCAAGAGTCTCTCAGGAGA 1148
DB |||||||
QY 2724 CCCAGGTCTCCCTGGCAAGATGGAGACCATGGAAACCTGGAAATTCAGGGCAACGAG 2783
DB |||||||
QY 1149 CCCAGGTCTCCCTGGCAAGATGGAGACCATGGAAACCTGGAAATTCAGGGCAACGAG 1208
DB |||||||
QY 2784 CCCCCAGGATCTCGGACCCATCACTATGTTTGTAGTGTAAATTCAGAGAGATTCGGT 2843
DB |||||||
QY 1209 CCCCCAGGATCTCGGACCCATCACTATGTTTGTAGTGTAAATTCAGAGAGATTCGGT 1268
DB |||||||
QY 2844 CAGAAAGACCAACTATTAG 2865
DB |||||||
QY 1269 CAGAAAGACCAACTATTAG 1290
DB |||||||
```

```
RESULT 2
US-10-037-270-282
; Sequence 282, Application US/10037270
; Publication No. US20030104529A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunging
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: Tillinghast, John
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. J320030104529A1el Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 784CIP2B
; CURRENT APPLICATION NUMBER: US/10/037,270
; CURRENT FILING DATE: 2002-01-04
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1104
; SOFTWARE: p7 FL_genes Version 1.0
; SEQ ID NO 282
; LENGTH: 2230
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCAT:CN: (85)...(1152)
US-10-037-270-282

Query Match 37.1%; Score 1062; DB 14; Length 2230;
Best Local Similarity 100.0%; Pred. No. 9.3e-284;
Matches 1062; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1804 GGAATCCCAAGGATTTCTCGAAACCCAGGATTAATGGGCCAAAGGAGAAATTCGGCCT 1863
DB |||||||
QY 91 GGAATCCCAAGGATTTCTCGAAACCCAGGATTAATGGGCCAAAGGAGAAATTCGGCCT 150
DB |||||||
QY 1864 CCAGGACAGCAAGGAAAAAAGGAGCCCGAGGATGCTCGTGGTAAATGGGAAGCAATGGC 1923
DB |||||||
QY 151 CCAGGACAGCAAGGAAAAAAGGAGCCCGAGGATGCTCGTGGTAAATGGGAAGCAATGGC 210
DB |||||||
QY 1924 TCACAGGCGCCCTCGAAACCCGGATCTAAGGGAAGCAAGGTGAACCTGGAATTCAA 1983
DB |||||||
QY 211 TCACAGGCGCCCTCGAAACCCGGATCTAAGGGAAGCAAGGTGAACCTGGAATTCAA 270
DB |||||||
QY 1984 GGGATGCTGGGGCTTCTGGGCTCAAGGAGAAACAGGAGCAACGGGTTCCCGAGGAA 2043
DB |||||||
QY 271 GGGATGCTGGGGCTTCTGGGCTCAAGGAGAAACAGGAGCAACGGGTTCCCGAGGAA 330
DB |||||||
QY 2044 CCAGGATACATGGTTTACCCGGGATTCAGGAAAAAAGGAGGGAACAAGGAAATCAAGGT 2103
DB |||||||
QY 331 CCAGGATACATGGTTTACCCGGGATTCAGGAAAAAAGGAGGGAACAAGGAAATCAAGGT 390
DB |||||||
QY 2104 GAAAGAGGTATTTCAGGTCAGGAGGAGAAATGGAAGACAGGAAATTCAGGGAACAG 2163
DB |||||||
QY 391 GAAAGAGGTATTTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTCAGG 450
DB |||||||
QY 2164 GGAATTCAGGTCATCATGGTGCAGAAAGGAGAGAGGTGAAAGGAGAACTCGGTGTC 2223
DB |||||||
QY 451 GGAATTCAGGTCATCATGGTGCAGAAAGGAGAGAGGTGAAAGGAGAACTCGGTGTC 510
DB |||||||
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2003, 19:13:33 ; Search time 702 Seconds  
(without alignments)

10410.575 Million cell updates/sec

Title: US-09-996-611B-5

Perfect score: 2865

Sequence: 1 atggctcactattacatt.....gaaagagcaaacatttag 2865

Scoring table: IDENTITY\_NUC

Gapop 10\_0 , Gapext 1.0

Searched: 1708419 seqs, 1275431651 residues

Total number of hits satisfying chosen parameters: 3416838

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA.\*

1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq.\*  
14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq.\*  
15: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
16: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*  
17: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	1282	44.7	2375	10	US-09-764-864-232
2	1062	37.1	2230	14	US-10-037-270-282
3	381	13.3	412	11	US-09-918-995-33525
4	285	9.9	4881	12	US-10-202-167-3
5	285	9.9	6352	12	US-10-202-167-1
6	267	9.3	496	11	US-09-918-995-17490
7	244.6	8.5	6545	14	US-10-198-846-13890
8	243	8.5	6281	14	US-10-084-817-155
9	242.6	8.5	5460	12	US-10-301-822-32
10	242.6	8.5	5460	14	US-10-177-293-67
11	242.6	8.5	5489	12	US-10-096-534-13
12	242.6	8.5	8664	14	US-10-198-846-9949
13	241.4	8.4	6319	12	US-10-096-534-15
14	239.8	8.4	6158	10	US-09-919-497-6
15	239.8	8.4	6158	10	US-09-954-456-762
16	239.8	8.4	6158	12	US-09-873-367C-432

17	233	8.1	4629	12	US-10-293-582-22	Sequence 22, Appl
18	233	8.1	4629	14	US-10-177-293-69	Sequence 69, Appl
19	233	8.1	6217	12	US-10-301-822-34	Sequence 34, Appl
20	231.8	8.1	4664	12	US-10-194-441A-2	Sequence 2, Appl
21	231.5	8.1	5060	14	US-10-175-523-181	Sequence 181, Appl
22	227.2	7.9	4660	12	US-10-133-013-132	Sequence 132, Appl
23	227.2	7.9	4660	14	US-10-084-817-154	Sequence 154, Appl
24	227.2	7.9	4837	14	US-10-084-817-45	Sequence 45, Appl
25	222.8	7.8	8664	14	US-10-198-846-9949	Sequence 9949, Ap
26	222.6	7.8	3671	14	US-10-084-817-67	Sequence 67, Appl
27	219.6	7.7	2511	14	US-10-157-031-61	Sequence 61, Appl
28	215	7.5	2542	10	US-09-964-824A-255	Sequence 255, Appl
29	215	7.5	2542	10	US-09-954-531-961	Sequence 961, Appl
30	212.8	7.4	6200	12	US-09-795-061-3	Sequence 3, Appl
31	211.4	7.4	6109	12	US-09-795-061-1	Sequence 1, Appl
32	207.8	7.3	7924	14	US-10-198-846-9752	Sequence 9752, Ap
33	207.6	7.2	5274	14	US-10-198-846-13840	Sequence 13840, A
34	206	7.2	5086	10	US-09-880-107-3947	Sequence 3947, Ap
35	206	7.2	5086	12	US-09-873-319-745	Sequence 745, App
36	206	7.2	5086	12	US-09-960-706-1123	Sequence 1123, Ap
37	206	7.2	5086	12	US-10-101-510-24	Sequence 24, Appl
38	206	7.2	5086	12	US-10-301-822-30	Sequence 30, Appl
39	206	7.2	5086	12	US-09-873-367C-522	Sequence 522, App
40	206	7.2	5086	12	US-09-873-367C-1067	Sequence 1067, Ap
41	206	7.2	5086	14	US-10-171-311-37	Sequence 37, Appl
42	206	7.2	5451	12	US-10-101-510-597	Sequence 597, App
43	205.8	7.2	4812	14	US-10-084-817-4	Sequence 4, Appl
44	205.8	7.2	5416	10	US-09-954-456-786	Sequence 786, App
45	205.8	7.2	5416	10	US-09-880-107-2094	Sequence 2094, Ap

#### ALIGNMENTS

RESULT 1  
US-09-764-864-232  
; Sequence 232, Application US/09764864  
; Patent No. US20020132753A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PT23  
; CURRENT APPLICATION NUMBER: US/09/764,864  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult: PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 1792  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 232  
; LENGTH: 2375  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-764-864-232

Query Match	44.7%	Score 1282:	DB 10;	Length 2375;
Best Local Similarity	100.0%	Pred. No. 0;		
Matches 1282;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1584	AAAGGTGAAATGGGTGCCAAAGGAGACAAAGGATCACCTGGATTATATGCAAAAGGG	1643	
DB	9	AAAGGTGAAATGGGTGCCAAAGGAGACAAAGGATCACCTGGATTATGCAAAAGGG	66	
QY	1640	TCCAAAGGTGAAAAGGGGATGCTGGCTTCCTGSCCTCCTGGACCTGCTGAGAAC	1703	
DB	69	TGCAAAAGGTGAAAAGGGGATGCTGGCTTCCTGSCCTCCTGGACCTGCTGAGAAC	128	
QY	1704	AGGAACATGAAAGGATGGATTATGGTAGTCCCGGTTCAGGGAGAGAGGAGATC	1763	
DB	129	AGGAACATGAAAGGATGGATTATGGTAGTCCCGGTTCAGGGAGAGAGGAGATC	188	
QY	1764	CCCTGTGCTCCGGSCAGGATGGAAACACGGGAGAGCTCCAGGATCCAGGATTCCTGG	1823	
DB	189	CCCTGTGCTCCGGSCAGGATGGAAACACGGGAGAGCTCCAGGATCCAGGATTCCTGG	246	

Job time : 10163 secs

c r

c r

62518	71007: contig of 8490 bp in length
71008	71107: gap of 100 bp
71108	79805: contig of 8698 bp in length
79806	79905: gap of 100 bp
81160	81160: contig of 1255 bp in length
81261	81260: gap of 100 bp
81261	92353: contig of 11093 bp in length
92354	92453: gap of 100 bp
92454	101135: contig of 8672 bp in length
101126	101235: gap of 100 bp
101226	117073: contig of 15848 bp in length
117074	117173: gap of 100 bp
117174	136020: contig of 18847 bp in length
136021	136120: gap of 100 bp
136121	161828: contig of 25708 bp in length
161829	161928: gap of 100 bp
161929	187999: contig of 26071 bp in length

FEATURES

```

/clone="RP11-18A3"
/clone_lib="RP01-11 Human Male BAC"
1..1227
/note="assembly_fragment"
1328..2416
/note="assembly_fragment"
2517..4057
/note="assembly_fragment"
4158..5171
/note="assembly_fragment"
5272..6380
/note="assembly_fragment"
6481..7659
/note="assembly_fragment"
7760..9214
/note="assembly_fragment"
9315..10932
/note="assembly_fragment"
11033..12035
/note="assembly_fragment"
12136..13215
/note="assembly_fragment"
13316..14645
/note="assembly_fragment"

```

```
Query Match          0.8%; Score 22; DB 2; Length 187999;
Best Local Similarity 100.0%; Pred. No. 7.6;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```



## SOURCE

Homo sapiens (human)

## ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 187999)

Birken, B., Linton, L., Nusbaum, C. and Lander, E.

Homo sapiens chromosome 1, clone RP11-18A3

Unpublished

2 (bases 1 to 187999)

Birken, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,

Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,

Boguslavsky, L., Boukagalter, B., Brown, A., Burkett, G.,

Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,

Collimore, A., Cooke, P., DeArrellano, K., Dewar, K., Diaz, J.S.,

Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D.,

Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,

Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,

Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,

Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lehoczky, J.,

Levine, R., Liu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,

McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,

Melgrim, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,

Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,

O'Neil, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,

Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,

Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,

Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,

Testaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,

Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,

Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission

Submitted (25-MAR-2000) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 187999)

Birken, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,

Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,

Boguslavsky, L., Boukagalter, B., Brown, A., Burkett, G.,

Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,

Collimore, A., Cooke, P., DeArrellano, K., Dewar, K., Diaz, J.S.,

Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D.,

Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,

Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,

Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,

Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lehoczky, J.,

Levine, R., Liu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,

McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,

Melgrim, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,

Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,

O'Neil, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,

Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,

Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,

Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,

Testaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,

Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,

Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission

Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Jul 22, 2000 this sequence version replaced gi:7328793.

All repeats were identified using RepeatMasker:

Smit, A.F.A. &amp; Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project information

Center project name: L7568

Center clone name: 18\_A\_3

----- Summary statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% cf reads

Assembly program: Phrap; version 0.96c731  
 Consensus quality: 167725 bases at least Q40  
 Consensus quality: 178025 bases at least Q30  
 Consensus quality: 182047 bases at least Q20  
 Insert size: 194000; agarose-fp  
 Insert size: 184399; sum-of-contigs  
 Quality coverage: 4.3 in Q20 bases; agarose-fp  
 Quality coverage: 4.5 in Q20 bases; sum-of-contigs

-----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 37 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

-----  
 \* 1 1227: contig of 1227 bp in length  
 \* 1228 1327: gap of 100 bp  
 \* 1328 2416: contig of 1089 bp in length  
 \* 2417 2516: gap of 100 bp  
 \* 2517 4057: contig of 1541 bp in length  
 \* 4058 4157: gap of 100 bp  
 \* 4158 5111: contig of 1014 bp in length  
 \* 5112 5271: gap of 100 bp  
 \* 5272 6380: contig of 1109 bp in length  
 \* 6381 7659: contig of 1179 bp in length  
 \* 7660 7759: gap of 100 bp  
 \* 7760 9214: contig of 1455 bp in length  
 \* 9215 9315: gap of 100 bp  
 \* 9316 10932: contig of 1618 bp in length  
 \* 10933 12035: contig of 1003 bp in length  
 \* 12036 12135: gap of 100 bp  
 \* 12136 13215: contig of 1080 bp in length  
 \* 13216 13315: gap of 100 bp  
 \* 13316 14645: contig of 1330 bp in length  
 \* 14646 14745: gap of 100 bp  
 \* 14746 15925: contig of 1180 bp in length  
 \* 15926 16025: gap of 100 bp  
 \* 16026 17158: contig of 1133 bp in length  
 \* 17159 17258: gap of 100 bp  
 \* 17259 19445: contig of 2187 bp in length  
 \* 19446 19545: gap of 100 bp  
 \* 19546 21242: contig of 1697 bp in length  
 \* 21243 21342: gap of 100 bp  
 \* 21343 24035: contig of 2693 bp in length  
 \* 24036 25142: contig of 1007 bp in length  
 \* 25143 25242: gap of 100 bp  
 \* 25243 27117: contig of 1875 bp in length  
 \* 27118 27217: gap of 100 bp  
 \* 27218 28912: contig of 1695 bp in length  
 \* 28913 29012: gap of 100 bp  
 \* 29013 31508: contig of 2496 bp in length  
 \* 31509 31608: gap of 100 bp  
 \* 31609 32697: contig of 1089 bp in length  
 \* 32698 32797: gap of 100 bp  
 \* 32798 37275: contig of 4478 bp in length  
 \* 37276 37375: gap of 100 bp  
 \* 37376 40952: contig of 3685 bp in length  
 \* 40953 41080: gap of 100 bp  
 \* 41081 45215: contig of 4135 bp in length  
 \* 45216 45315: gap of 100 bp  
 \* 45316 48710: contig of 3395 bp in length  
 \* 48711 48810: gap of 100 bp  
 \* 48811 52850: contig of 4340 bp in length  
 \* 52851 52950: gap of 100 bp  
 \* 52951 57072: contig of 4122 bp in length  
 \* 57073 57172: gap of 100 bp  
 \* 57173 62417: contig of 5245 bp in length  
 \* 62418 62517: gap of 100 bp

```

misc_feature      82998..122831
                  /note="assembly_name:Contig11
                  clone_end:T7
                  vector_side:right"
misc_feature      122932..176694
                  /note="assembly_name:Contig12"
BASE COUNT      48989 a 36992 c 36355 g 53458 t      900 others
ORIGIN
Query Match      0.8%; Score 22; DB 2; Length 176694;
Best Local Similarity 100.0%; Pred.No. 7.7;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1862 CTCAGGACAGCAGGAGGAAAAA 1883
        |||
Db      135116 CTCAGGACAGCAGGAGGAAAAA 135137

RESULT 42
AL663099/c
LOCUS      AL663099      177135 bp      DNA      linear      ROD 19-JUL-2002
DEFINITION Mouse DNA sequence from clone DN-384L23 on chromosome 3, complete
sequence.
ACCESSION      AL663099.7 GI:23068691
VERSION      AL663099.7
KEYWORDS      HTG.
SOURCE      Mus musculus (house mouse)
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE      1 (bases 1 to 177135)
AUTHORS      Clark,G.
TITLE      Direct Submission
JOURNAL      Submitted (18-JUL-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries: clonerequests@sanger.ac.uk
humquery@sanger.ac.uk
On Apr 7, 2002 this sequence version replaced gi:20067508.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em., EMBL; Sw.,
SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
-----
DN-384L23 is from a DIL (Diabetes and Inflammation Laboratory) NCD
Mouse BAC library
VECTOR: pBACe3.6.
-----
FEATURES
source
1. 177135
   /organism="Mus musculus"
   /mol_type="genomic DNA"
   /db_xref="taxon:10090"
   /chromosome="3"
   /clone="DN-384L23"
   /clone_lib="NCD mouse library"
BASE COUNT      52554 a 40202 c 38941 g 45438 t

```

```

ORIGIN
Query Match      0.8%; Score 22; DB 10; Length 177135;
Best Local Similarity 100.0%; Pred.No. 7.7;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      23480 GAGAGTCTTCAGACAAATTAT 2369
        |||
Db      36082 GAGAGTCTTCAGACAAATTAT 36061

RESULT 43
AC122397
LOCUS      AC122397      181742 bp      DNA      linear      ROD 02-MAR-2003
DEFINITION Mus musculus chromosome 15 clone RP24-90D21, complete sequence.
ACCESSION      AC122397
VERSION      AC122397.2 GI:28630148
KEYWORDS      HTG.
SOURCE      Mus musculus (house mouse)
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE      1 (bases 1 to 181742)
AUTHORS      McPherson,J.D. and Waterston,R.H.
TITLE      The sequence of Mus musculus clone
JOURNAL      Unpublished
REFERENCE      2 (bases 1 to 181742)
AUTHORS      McPherson,J.D. and Waterston,R.H.
TITLE      Direct Submission
JOURNAL      Submitted (23-MAY-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE      3 (bases 1 to 181742)
AUTHORS      McPherson,J.D. and Waterston,R.H.
TITLE      Direct Submission
JOURNAL      Submitted (02-MAR-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
COMMENT      Or Mar 2, 2003 this sequence version replaced gi:21105847.
-----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submiss@wustl.wustl.edu
-----
Center project name: M.BB0090D21
-----
FEATURES
source
1. 181742
   /organism="Mus musculus"
   /mol_type="genomic DNA"
   /db_xref="taxon:10090"
   /chromosome="5"
   /clone="RP24-90D21"
BASE COUNT      48193 a 43416 c 42696 g 47437 t
ORIGIN
Query Match      0.8%; Score 22; DB 10; Length 181742;
Best Local Similarity 100.0%; Pred.No. 7.6;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1507 GGAGAACCGAGCGGAGATGGTG 1528
        |||
Db      21276 GGAGAACCGAGCGGAGATGGTG 21297

RESULT 44
AC026928/c
LOCUS      AC026928      187999 bp      DNA      linear      HTG 24-AUG-2002
DEFINITION Homo sapiens chromosome 1 clone Rp11-18A3 map 1, WORKING DRAFT
SEQUENCE, 37 unordered pieces.
ACCESSION      AC026928
VERSION      AC026928.2 GI:9369514
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT.

```

```
-----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
-----
Project Information
-----
Center project name: H.NR0237E05
-----
Summary Statistics
-----
```

Sequencing vector: M13; 1008

Sequencing vector: plasmid; 0%  
Chemistry: Dye-primer ET; 100% of reads  
Assembly: Dye-terminator Big Dye; % of reads  
Assembly program: Phrap; version 0.95c3.9  
Consensus quality: 172839 bases at least Q40  
Consensus quality: 17357 bases at least Q30  
Consensus quality: 174626 bases at least Q20  
Insert size: 177050; agarose-fp  
Insert size: 175794; sum-of-contigs  
Quality coverage: 5.63 in Q20 bases; aggregate  
Quality coverage: 5.69 in Q20 bases; sum-of-contigs

\* NOTE: this is a 'working draft' sequence - it currently  
\* consists of 10 configs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the configs are represented as  
\* runs of 'N', but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1	1659:	contig	of 1659 bp in length
1660	1759:	gap	of unknown length
1760	6808:	contig	of 5049 bp in length
6809	6908:	gap	of unknown length
6909	13807:	contig	of 6899 bp in length
13808	13927:	gap	of unknown length
20324	20324:	contig	of 6417 bp in length
20424	20424:	gap	of unknown length
20425	27826:	contig	of 7402 bp in length
27827	27926:	gap	of unknown length
27927	36369:	contig	of 8443 bp in length
36370	36469:	gap	of unknown length
50727	50727:	contig	of 14258 bp in length
50827	50827:	gap	of unknown length
82897	82897:	contig	of 32070 bp in length
82997	82997:	gap	of unknown length
122831	122831:	contig	of 39834 bp in length
122931	122931:	gap	of unknown length
176694	176694:	contig	of 53763 bp in length

## Location/Qualifiers

```
location/724411115
source
1..176694
/organism:"Homo sapiens"
/mol_type:"genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/clone:"RP11-237E5"
1..1655
/misc_feature
/note="assembly_name:Contig3"
1760..6898
/misc_feature
/note="assembly_name:Contig4"
6909..13907
/misc_feature
/note="assembly_name:Contig5"
13908..20324
/misc_feature
/note="assembly_name:Contig6"
20425..27826
/misc_feature
/note="assembly_name:Contig7"
27947..36369
/misc_feature
/note="assembly_name:Contig8"
clone_end:SP6
vector_side:left"
36470..50727
/misc_feature
/note="assembly_name:Contig9"
50828..82897
/misc_feature
/note="assembly_name:Contig10"
```

DS  
C:

JUL-2023  
EQUENCE,

stori;

n . Louis,

```

misc_feature 497. 543
/notes="single clone coverage"

misc_feature 14367. 14393
/notes="single clone coverage"

misc_feature 47455. 47504
/notes="single clone coverage"

misc_feature 54353. 54525
/notes="single clone coverage"

misc_feature 55101. 55135
/notes="single clone coverage"

misc_feature 63051.63052
/notes="bacterial transposon excised; IS2 sequence can be
found in GenBank Accession Number AF300143.1 nucleotides
1305-2635; 1331 bp"

misc_feature 76511. 76572
/notes="single clone coverage"

misc_feature 98071. 98107
/notes="single clone coverage"

misc_feature 130697. 130699
/notes="single clone coverage"

misc_feature 151137. 151299
/notes="single clone coverage"

misc_feature 151338. 151471
/notes="single clone coverage"

misc_feature 163358. 163377
/notes="single clone coverage"

misc_feature 163419. 163469
/notes="single clone coverage"

BASE COUNT 47092 a 33553 c 35099 g 51007 t
ORIGIN

Query Match 0.8%; Score 22; DB 10; Length 166751;
Best Local Similarity 100.0%; Pred. No. 7.7;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1183 AAAGAGAAACTGTTTCAGTTG 1204
|||||
Db 6429 AAAGAGAAACTGTTTCAGTTG 6450

RESULT 40
AC102838/c
LOCUS
DEFINITION AC102838 169947 bp DNA linear HTG 23-MAR-2003
Mus musculus clone RP24-317M2, WORKING DRAFT SEQUENCE, 9 unordered
pieces.
ACCESSION AC102838
VERSION GI:29164563
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 169947)
Birren,B., Nusbaum,C. and Lander,E.
Unpublished
Mus musculus, clone RP24-317M2
2 (bases 1 to 169947)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Coepele,V., Colangelo,M., Collins,S., Collumore,A., Cooke,A.,
Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Govette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., Larocque,K.,
Lamarez,R., Landers,T., Lehocsky,J., Levine,R., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,

```

```

Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 169947)
Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
Boguslavsky,L., Boukhgalter,B., Camarata,J., Chang,J., Choepele,Y.,
Collumore,A., Cooke,A., Cooke,P., Cotum,B., DeArelano,K.,
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
Graham,L., Grand-Pierre,N., Hafez,N., Hagepiat,D., Hagos,B.,
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
Meldrim,C., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schupack,R., Seaman,S., Severy,P., Smith,C.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (23-MAR-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 23, 2003 this sequence version replaced gi:22381848.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIER
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: 119940
Center clone name: 317 M.2
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 166683 bases at least Q40
Consensus quality: 167466 bases at least Q30
Consensus quality: 167808 bases at least Q20
Insert size: 173000; agarose-fp
Insert size: 169147; sum-of-contigs
Quality coverage: 11.7 in Q20 bases; agarose-fp
Quality coverage: 12.0 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 35676: contig of 35676 bp in length
* 35677 35776: gap of 100 bp
* 35777 36492: contig of 716 bp in length
* 36493 36592: gap of 100 bp
* 36593 37265: contig of 673 bp in length
* 37266 37365: gap of 100 bp
* 37366 38939: contig of 1574 bp in length
* 38940 39039: gap of 100 bp

```

TITLE  
JOURNAL  
AUTHORS  
REFERENCE

TITLE  
JOURNAL  
AUTHORS  
REFERENCE

COMMENT

```

/misc_feature
  /chromosome="12"
  /clones="RP11-22B23"
  1..2004
  /note="overlaps bases 21010..23013 of clone AC137766"
  2636..2726
  /rpt_family="L2"
  3337..3642
  /rpt_family="MER2"
  3927..4028
  /rpt_family="L2"
  complement(4477..4535)
  /rpt_family="L2"
  complement(5459..5982)
  /rpt_family="MLTID"
  complement(5906..6036)
  /rpt_family="FLAM_A"
  6077..6906
  /rpt_family="L1ME1"
  6932..7036
  /rpt_family="L1ME3"
  7047..7736
  /rpt_family="L1ME3"
  7872..7913
  /rpt_family="AT_rich"
  7939..8120
  /rpt_family="L1ME3"
  8121..8414
  /rpt_family="AluY"
  8415..8541
  /rpt_family="L1ME3"
  8542..8729
  /rpt_family="AluJo"
  8786..9242
  /rpt_family="L1ME3"
  10126..10620
  /rpt_family="L1P13"
  complement(11724..11881)
  /rpt_family="MIR"
  13186..13271
  /rpt_family="MIR"
  complement(13276..13406)
  /rpt_family="MIR"
  complement(14360..14818)
  /rpt_family="LTR36"
  complement(16338..16431)
  /rpt_family="L2"
  16785..16978
  /rpt_family="MER101"
  16979..17423
  /rpt_family="LTR46"
  17424..17488
  /rpt_family="MER101"
  complement(17600..17702)
  /rpt_family="L2"
  18161..18264
  /rpt_family="MIR"
  18459..18561
  /rpt_family="MIR"
  18614..18894
  /rpt_family="AluDb"
  18895..18931
  /rpt_family="(TAAA)n"
  18932..18944
  /rpt_family="AluDb"
  complement(20133..20222)
  /rpt_family="AluDb"
  Query_Match
  Best Local Similarity 100.0%; Pred.No. 7.7;
  Matches 22; Conservative 0; Mismatches 0; Indels 0; Caps 0;
  Length 162572;
  0.8%; Score 22; DB 9;

```

```

Db 140464 CTGGAAGAGAACTGTTC A140485

RESULT 39
AC087066
LOCUS
DEFINITION
  Rattus norvegicus clone RP31-194D8 strain Brown Norway, complete
  sequence.
ACCESSION
AC087066
VERSION
AC087066.3 GI:14277283
KEYWORDS
  RTG.
SOURCE
  Rattus norvegicus (Norway rat)
  ORGANISM
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
    Rattus.
  REFERENCE
    1 (bases 1 to 166751)
      Auye, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W.,
      Bouffard, G.G., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S.,
      Guan, X., Gupta, J., Ho, S.-L., Idol, J.R., Karlins, E., Lee-Lin, S.-Q.,
      Legaspi, R., Lim, M., Maduro, Q.L., Maduro, V.B., Masiello, C.,
      Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Prasad, A.,
      Shevchenko, Y., Snyder, B., Stantripop, S., Thomas, J.W., Thomas, P.J.,
      Tjongson, E.E., Touchman, J.W., Tsugeon, C., Vogt, J.L., Walker, M.A.,
      Wetherby, K.J., Zhang, L.-H. and Green, E.D.
      NISC Comparative Sequencing Initiative
      Unpublished
      2 (bases 1 to 166751)
      Green, E.D.
      Direct Submission
      Submitted (05-DEC-2000) NIH Intramural Sequencing Center, 8717
      Grovermont Circle, Gaithersburg, MD 20877, USA
      REFERENCE
      3 (bases 1 to 166751)
      Green, E.D.
      Direct Submission
      Submitted (17-MAY-2001) NIH Intramural Sequencing Center, 8717
      Grovermont Circle, Gaithersburg, MD 20877, USA
      REFERENCE
      4 (bases 1 to 166751)
      Green, E.D.
      Direct Submission
      Submitted (01-JUN-2001) NIH Intramural Sequencing Center, 8717
      Grovermont Circle, Gaithersburg, MD 20877, USA
      On Jun 1, 2001 this sequence version replaced gi:14140275.
      ----- Genom Center
      Center: NIH Intramural Sequencing Center
      Center code: NISC
      Web site: http://www.nisc.nih.gov
      Contact: nisc.mouses@hgrl.nih.gov
      ----- Project Information
      Center project name: ahn
      Center clone name: 194D08

```

This sequence was finished as follows unless otherwise noted:  
 all regions were double-stranded, sequenced with an  
 alternate chemistry, or covered by high quality data  
 (i.e., paired quality >= 30); an attempt was made to resolve  
 all sequencing problems, such as compressions and repeats;  
 all regions were covered by at least one plasmid subclone  
 or more than one M13 subclone; and the assembly was confirmed  
 by restriction digest.

CLONE LENGTH: This sequence represents the entire insert of  
 this clone unless otherwise noted. If there are overlapping  
 clones, the overlaps are noted in the beginning and end of  
 the Features section.

```

FEATURES
  Location/Qualifiers
    1..166751
      /organism="Rattus norvegicus"
      /mol_type="genomic DNA"
      /strain="Brown Norway"
      /db_xref="taxon:10116"
      /clone="RP31-194D8"
      /clone_lib="RP31"
  source

```

Query Match 0.8%; Score 22; DB 2; Length 157793;

Best Local Similarity 100.0%; Pred. NO. 7.7;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2348 GAGAGTTTTTCAGACAATTAT 2369

Db 125386 GAGAGTTTTTCAGACAATTAT 125407

# RESULT 38

AC009533

LOCUS

DEFINITION Homo sapiens 12 BAC RP11-22B23 (Roswell Park Cancer Institute Human

BAC Library) complete sequence.

ACCESSION AC009533

VERSION AC009533.10

KEYWORDS GI:27436734

SOURCE HTG.

ORGANISM Homo sapiens (human)

REFERENCE

AUTHORS

1 (bases 1 to 162572)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 Musny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,  
 Alsorooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Barks,T.,  
 Barbara,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,  
 Bouck,J., Bowie,S., Briefe,M., Brown,E., Brown,M., Bryant,M.P.,  
 Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,  
 Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,  
 Chen,G., Chen,R., Chen,Z., Chiu,D., Chowdhry,I., Christopoulos,C.,  
 Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,  
 Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,  
 Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,  
 Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,  
 Earnhart,C., Edgar,D., Edwards,C.C., Eihaj,C., Emerling,S.,  
 Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P.,  
 Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,C., Garza,N.,  
 Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,  
 Hamilton,K., Han,J., Harris,C., Hartis,K., Hart,M., Havlak,P.,  
 Hayes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogue,M.,  
 Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S.,  
 Hume,J., Ioshikhes,I., Jackson,L.E., Jacobson,B., Jia,Y.,  
 Johnson,R., Jolivet,S., Joudan,S., Karlsson,E., Kelly,S., Khan,J.,  
 King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N.,  
 Leal,B., Lee,E., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O.,  
 Lieu,C., Liu,J., Liu,M., Louseged,H., Lozado,R.J., Lu,X.,  
 Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P.,  
 Marandel,I., Martin,R., Martindale,A., Martinez,E., Massey,E.,  
 Mawhinney,E., McLeod,M.P., Meador,M., Mei,G., Merscher,S.,  
 Metzker,M., Miller,A., Miner,G., Miner,Z., Mitchell,T.,  
 Mohabbat,K., Montgomery,K.T., Morgan,M., Morris,S., Moser,M.,  
 Neal,D., Nelson,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,  
 Nguyen,N., Nickerson,E., Nwokenwo,S., Oguh,M., Okwuonu,G.,  
 Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,  
 Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,  
 Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G.,  
 Scherer,S., Scott,G., Shen,H., Shim,C., Shooshari,N., Sisson,I.,  
 Sodergren,E., Sonaikhe,T., Sparks,A., Stanley,H., Stone,H.,  
 Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,  
 Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,  
 Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,O.,  
 Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,  
 Williams,G., Williamson,A., Wleciyk,R., Wooden,S., Worley,K.,  
 Wu,C., Wu,Y.F., Zhou,J., Zorrilla,S., Kucherlapati,R.,  
 Weinstock,G. and Gibbs,R.

Direct Submission

Unpublished

2 (bases 1 to 162572)

Worley,K.C.

Direct Submission

Submitted (27-AUG-1999) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA

REFERENCE  
 AUTHORS

TITLE  
 JOURNAL

REFERENCE  
 AUTHORS

TITLE  
 JOURNAL

REFERENCE  
 AUTHORS

TITLE  
 JOURNAL

REFERENCE  
 AUTHORS

TITLE  
 JOURNAL

REFERENCE  
 AUTHORS

TITLE  
 JOURNAL

COMMENT

3 (bases 1 to 162572)

Worley,K.C.

Direct Submission

Submitted (31-DEC-1999) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA

4 (bases 1 to 162572)

Worley,K.C.

Direct Submission

Submitted (05-JAN-2000) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA

5 (bases 1 to 162572)

Worley,K.C.

Direct Submission

Submitted (16-MAY-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA

6 (bases 1 to 162572)

Worley,K.C.

Direct Submission

Submitted (31-DEC-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA

7 (bases 1 to 162572)

Worley,K.C.

Direct Submission

Submitted (08-JAN-2003) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA

On Dec 31, 2002 this sequence version replaced gi:66499266.

INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email

gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the  
 entire insert of this clone. Overlapping regions of clones are only  
 sequenced and submitted once, so the sequence for the remainder of  
 the insert may be found in the record for the adjacent clones.  
 Overlapping clones are noted at the beginning and end of the  
 Features listing.

## ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches  
 of a local database that includes entries from dbSTS, GDB, and  
 local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green,  
 unpublished) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST  
 (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the  
 EST and cDNA sequences. Genes demonstrate at least two exons  
 flanked by consensus splice sites that maintained sequence  
 continuity across the splice junctions. Sequences that are not  
 identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum  
 standard of double strand coverage with a minimum of 2 clones and 2  
 reads with no ambiguities or 2 chemistries with a minimum of 2  
 clones and 3 reads with no ambiguities. If the sequence quality for  
 a region does not meet this standard, it will be indicated in the  
 annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality  
 standards - estimated error rate less than 1 per 10,000 bases.  
 Reports of lowest quality individual bases and measures of base  
 quality are listed below. Description of the metrics can be found  
 at URL:

<http://www.hgsc.bcm.tmc.edu:8088/quality.infc/genbank.annotation.ht>

Location/Qualifiers

1..162572

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

FEATURES

source

RESULT 37

```

AC102109          157793 bp      DNA      linear      HTG 12-JUN-2003
LOCUS             Mus musculus clone RP23-113117, WORKING DRAFT SEQUENCE, 9 unordered
DEFINITION        pieces.
ACCESSION         AC102109.3 GI:31621376
VERSION           HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.
KEYWORDS          Mus musculus (house mouse)
SOURCE            Mus musculus
ORGANISM          Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE         1 (bases 1 to 157793)
AUTHORS           Birren B., Nusbaum C. and Lander E.
JOURNAL           Mus musculus, clone RP23-113117
TITLE             Unpublished
REFERENCE         2 (bases 1 to 157793)
AUTHORS           Birren B., Linton L., Nusbaum C., Lander E., Ali A., Allen N.,
Anderson S., Barna N., Bastien V., Boguslavskiy L., Boukhgalter B.,
Brown A., Camarata J., Campopiano A., Chang J., Chazaro B.,
Choepe Y., Collangelo M., Collins S., Collymore A., Cook A.,
Cooke P., Dearellano K., Dewar K., Diaz J.S., Dodge S., Faro S.,
Ferreira P., Fitzhugh W., Gage D., Galagan J., Gardyna S.,
Gande S., Gord S., Goyette M., Graham L., Grand-Pierre N.,
Hagos B., Heaford A., Horton L., Huime W., Kells C., LaRoque K.,
Jones C., Kamat A., Karatas A., Lehoucq J., Levine R., Liu G.,
Lamazares R., Landers T., Lehoucq J., Marquis N., Matthews C.,
Maclean C., Macdonald P., Major J., Marquis N., Matthews C.,
McCarthy M., McEwan P., McKernan K., MCPheeters R., Moulden J.,
Meneus L., Mihova T., Mlenga V., Murphy T., Naylor J., Nguyen C.,
Norbu C., Norman C.H., O'Connor T., O'Donnell P., O'Neill D.,
Oliver J., Peterson K., Phunkhang P., Pierre N., Pollara V.,
Raymond C., Retta R., Ribback M., Riley R., Rise C., Rogov P.,
Roman J., Rosetti M., Roy A., Santos R., Schauer S., Schupback R.,
Seaman S., Severy P., Spencer B., Stange-Thomann N., Stojanovic N.,
Strauss N., Subramanian A., Talamas J., Tesfaye S., Theodore J.,
Topham K., Travers M., Travis N., Trigglio J., Vassiliev H.,
Viel R., Vo A., Wilson B., Wu X., Wyman D., Ye W.J., Young G.,
Zainoun J., Zembek L., Zimmer A. and Zody M.
JOURNAL           Direct Submission
TITLE             Submitted (23-NOV-2003) Whitehead Institute/MIT Center for Genome
JOURNAL           Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE         3 (bases 1 to 157793)
AUTHORS           Birren B., Nusbaum C., Lander E., Abouelleil A., Allen N.,
Anderson S., Arachchi H.M., Barna N., Bastien V., Blood T.,
Boguslavskiy L., Boukhgalter B., Camarata J., Chang J., Choepe Y.,
Collamore A., Cook A., Cooke P., Corum B., Dearellano K.,
Diaz J.S., Dodge S., Dooley K., Dorris L., Erickson J., Farc S.,
Ferreira P., FitzGerald M., Gage D., Galagan J., Gardyna S.,
Graham L., Grand-Pierre N., Hafez N., Hagopian D., Hagos B.,
Hall J., Horton L., Huime W., Iliev I., Johnson R., Jones C.,
Kamat A., Karatas A., Kells C., Landers T., Levine R.,
Lindblad-Toh K., Liu G., Lui A., Mabbitt R., Maclean C.,
Macdonald P., Major J., Manning J., Matthews C., McCarthy M.,
Meldrim J., Meneus L., Mihova T., Mlenga V., Murphy T., Naylor J.,
Nguyen C., Nicol R., Norbu C., O'Connor T., O'Donnell P.,
O'Neill D., Oliver J., Peterson K., Phunkhang P., Pierre N.,
Rachupka A., Ramasamy U., Raymond C., Retta R., Rise C., Rogov P.,
Roman J., Schauer S., Schupback R., Seaman S., Severy P., Smith C.,
Spencer B., Stange-Thomann N., Stojanovic N., Stubbs M.,
Talamas J., Tesfaye S., Theodore J., Topham K., Travers M.,
Vassiliev H., Venkataraman V.S., Viel R., Vo A., Wilson B., Wu X.,
Wyman D., Young G., Zainoun J., Zembek L., Zimmer A. and Zody M.
JOURNAL           Direct Submission
TITLE             Submitted (12-JUN-2003) Whitehead Institute/MIT Center for Genome
JOURNAL           Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT           On Jun 12, 2003 this sequence version replaced gi:22381255.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RX/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WBIR

```

Web site: <http://www.seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information  
Center project name: L18049  
Center clone name: 113-117  
----- Summary Statistics  
Sequencing vector: Plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% cf reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 155596 bases at least Q43  
Consensus quality: 156505 bases at least Q30  
Consensus quality: 156946 bases at least Q20  
Insert size: 182000; agarose-fp  
Insert size: 156993; sum-of-ctg  
Quality coverage: 10.7 in Q20 bases; agarose-fp  
Quality coverage: 12.4 in Q20 bases; sum-of-ctg

-----  
\* NOTE: This is a 'working draft' sequence. It currently  
consists of 9 contigs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as  
runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence.  
\* as soon as it is available and the accession number will  
be preserved.

\* 40918: cntig of 40918 bp in length  
\* 40919: gap of 100 bp  
\* 41019: 41663: cntig of 645 bp in length  
\* 41664: 41763: gap of 100 bp  
\* 41764: 43522: cntig of 1759 bp in length  
\* 43523: 43622: gap of 100 bp  
\* 43623: 44624: cntig of 1002 bp in length  
\* 44625: 44724: gap of 100 bp  
\* 44725: 47305: cntig of 2581 bp in length  
\* 47306: 47405: gap of 100 bp  
\* 47406: 61096: cntig of 13691 bp in length  
\* 61097: 81339: gap of 100 bp  
\* 81339: 81439: cntig of 20343 bp in length  
\* 81440: 81439: gap of 100 bp  
\* 11243: 112412: cntig of 30973 bp in length  
\* 112513: 112512: gap of 100 bp  
\* 112513: 157793: cntig of 45201 bp in length.

FEATURES  
source

1..157793  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/clone="RP23-113117"  
/clone\_lib="RP23-113117 Female Mouse BAC"

misc\_feature

1..40918  
/notes="assembly\_fragment"  
clone\_end:SP6  
vector\_side:left

misc\_feature

41019..41663  
/notes="assembly\_fragment"

misc\_feature

41764..43522  
/notes="assembly\_fragment"

misc\_feature

43623..44624  
/notes="assembly\_fragment"

misc\_feature

44725..47305  
/notes="assembly\_fragment"

misc\_feature

47406..61096  
/notes="assembly\_fragment"

misc\_feature

61197..81339  
/notes="assembly\_fragment"

misc\_feature

81440..112412  
/notes="assembly\_fragment"

misc\_feature

112513..157793  
/notes="assembly\_fragment"

misc\_feature

clone\_end:T7  
vector\_side:right

BASE COUNT 42452 a 34126 c 35298 g 45105 t 812 others  
ORIGIN

```
/chromosome="1"
/clone="RP11-183G22"
/clone_lib="RPC1-11.1"
96048..96468
/note="Sequence from clone PCR only."
misc_feature
96070..96096
/note="Sequence from uni-directional dGTP big dye
terminator reads only."
BASE COUNT 41262 a 31518 c 29234 g 37259 t
ORIGIN
Query Match 0.8%; Score 22; DB 9; Length 139273;
Best Local Similarity 100.0%; Pred. No. 7.8;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1175 ATTCTGGAAGAAGAACTGT 1196
|||||
DB 138156 ATTCTGGAAGAAGAACTGT 138177

RESULT 35
AL928839 140119 bp DNA linear ROD 21-DEC-2002
LOCUS Mouse DNA sequence from clone RP23-298M2 on chromosome 2, complete
sequence.
ACCESSION AL928839
VERSION AL928839.10 GI:27368337
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Griffiths, C.
REFERENCE 1 (bases 1 to 140119)
AUTHORS Direct Submission
TITLE Submitted (20-DEC-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonequest@sanger.ac.uk
JOURNAL On Dec 23, 2002 this sequence version replaced gi:27262731.
COMMENT Sequence from the Mouse Genome Sequencing Consortium whole genome
shotgun may have been used to confirm this sequence. Sequence data
from the whole genome shotgun alone has only been used where it has
a phred quality of at least 30.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
-----
```

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Mp., WORMPEP; Information on the WORMPEP database can be found at <http://www.sanger.ac.uk/Projects/C.elegans/wormpep> RP23-298M2 is from the RPC1-23 Mouse PAC Library

constructed by the group of Pieter de Jong.  
For further details see <http://www.chori.org/bacpac/home.htm>  
VECTOR: pBACe3.6.

FEATURES Location/Qualifiers

```
source
1..140119
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="2"
/clone="RP23-298M2"
/clone_lib="RPC1-23"
BASE COUNT 42768 a 25589 c 25793 g 45969 t
ORIGIN
Query Match 0.8%; Score 22; DB 10; Length 140119;
Best Local Similarity 100.0%; Pred. No. 7.8;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 858 TCAAGATTAAAGTCAAGAAA 879
|||||
DB 114497 TCAAGATTAAAGTCAAGAAA 114518

RESULT 36
AC122459 156807 bp DNA linear RCD 19-OCT-2002
LOCUS Mus musculus chromosome 15 clone RP24-273F21, complete sequence.
DEFINITION AC122459
ACCESSION AC122459
VERSION AC122459.2 GI:24137602
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
McPherson, J.D. and Waterston, R.H.
REFERENCE 1 (bases 1 to 156807)
AUTHORS The sequence of Mus musculus clone
TITLE Unpublished
JOURNAL
REFERENCE 2 (bases 1 to 156807)
AUTHORS McPherson, J.D. and Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (23-MAY-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 3 (bases 1 to 156807)
AUTHORS McPherson, J.D. and Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (19-OCT-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
COMMENT On Oct 19, 2002 this sequence version replaced gi:21155920.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.wustl.edu
----- Project Information
Center project name: M B30273F21
----- Location/Qualifiers
source
1..156807
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="15"
/clone="RP24-273F21"
BASE COUNT 42969 a 35824 c 36400 g 41614 t
ORIGIN
Query Match 0.8%; Score 22; DB 10; Length 156807;
Best Local Similarity 100.0%; Pred. No. 7.7;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1507 GGAGAACCGCGCGAGATGGT 1528
|||||
DB 153405 GGAGAACCGCGCGAGATGGT 153429
```



```

repeat_region complement(3519..3717)
repeat_region /rpt_family="AluSg/x"
repeat_region 5701..5804
repeat_region /rpt_family="MER5A"
repeat_region 5873..5971
repeat_region /rpt_family="L1M4"
repeat_region 7397..7559
repeat_region /rpt_family="MER20"
repeat_region 8700..8736
repeat_region /rpt_family="(TTG)n"
repeat_region 9223..9258
repeat_region /rpt_family="U2"
repeat_region 10121..10146
repeat_region /rpt_family="(TTTA)n"
repeat_region complement(10158..10386)
repeat_region /rpt_family="L1PA10"
repeat_region complement(11193..11473)
repeat_region 11954..11993
repeat_region /rpt_family="AluJo"
repeat_region 12497..12631
repeat_region /rpt_family="(T)n"
repeat_region 12633..12815
repeat_region /rpt_family="MIR"
repeat_region 12954..13095
repeat_region /rpt_family="MSTC"
repeat_region 13871..14150
repeat_region /rpt_family="AluY"
repeat_region 14380..14563
repeat_region /rpt_family="MIR"
repeat_region complement(15192..15480)
repeat_region /rpt_family="AluSx"
repeat_region 15613..15671
repeat_region /rpt_family="MIR"
repeat_region complement(15994..16243)
repeat_region 16413..16688
repeat_region /rpt_family="AluY"
repeat_region 17979..18000
repeat_region /rpt_family="AT rich"
repeat_region complement(20594..20764)
repeat_region /rpt_family="L1PA12"
STS 21925..22322
/standard_name="PZF_F/PZF_R"
/db_xref="GDB:214828"
22447..22548
/rpt_family="MIR"
23228..23686
/rpt_family="L1PA10"
repeat_region 30266..30792
/rpt_family="MLTIF"
repeat_region complement(31334..32052)
/rpt_family="MER44C"
32404..32529
/standard_name="DL251257"
/db_xref="DBSTS:8833"
32536..32640
/rpt_family="L2"
33140..33305
/rpt_family="MIR"
33343..33364
/rpt_family="(TTTA)n"
repeat_region complement(33365..33444)
/rpt_family="L1PA2"
33445..34137
/rpt_family="L1p"
repeat_region 35050..35147
/rpt_family="MIR"
repeat_region 36398..36627
/rpt_family="L1M9A"
repeat_region complement(36628..36907)
/rpt_family="AluJo"
37345..39018

```

```

/rpt_family="L1PA13"
complement(39019..39324)
/rpt_family="AluSc"
39325..39818
/rpt_family="L1PA13"
repeat_region 39819..40218

Query Match      0.8%; Score 22; DB 9; Length 127277;
Best Local Similarity 100.0%; Pred.No.7.9;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1178 CTGAAAGAGAACTGTTC A 1199
Db 94338 CTGAAAGAGAACTGTTC A 94315

RESULT 34
AL596275
LOCUS
DEFINITION
Humar DNA sequence from clone Rp11-183G22 on chromosome 1, complete
sequence.
AL596275
VERSION
AL596275.10 GI:17384503
KEYWORDS
HTG.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 139273)
REFERENCE
Lawlor,S.
Direct Submission
Submitted (04-DEC-2001) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Dec 5, 2001 this sequence version replaced gi:15973968.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30; an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: EMI, EMBL; SW,
SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
This sequence
was generated from part of bacterial clone contigs of human
chromosome 1, constructed by the Sanger Centre Chromosome Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/chr1
Rp11-183G22 is from the library RPI-11.1 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6
IMPORTANT: This sequence is not the entire insert of clone
Rp11-183G22. It may be shorter because we sequence overlapping
sections only once, except for a short overlap.
The true left end of clone Rp11-183G22 is at 1 in this sequence.
The true left end of clone RP5-850015 is at 137274 in this
sequence. The true right end of clone Rp11-567C20 is at 88255 in
this sequence.

FEATURES
Location/Qualifiers
1..139273
/crganism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

```







```

* 41340 41439: gap of 100 bp
* 41440 42110: contig of 671 bp in length
* 42111 42120: gap of 100 bp
* 42121 42863: contig of 653 bp in length
* 42864 42964: contig of 100 bp
* 42965 43614: contig of 651 bp in length
* 43615 43714: gap of 100 bp
* 43715 44362: contig of 648 bp in length
* 44363 44462: gap of 100 bp
* 44463 45008: contig of 546 bp in length
* 45009 45108: gap of 100 bp
* 45109 45756: contig of 648 bp in length
* 45757 45856: gap of 100 bp
* 45857 46512: contig of 656 bp in length
* 46513 46612: gap of 100 bp
* 46613 47287: contig of 675 bp in length
* 47288 47387: gap of 100 bp
* 47388 48042: contig of 655 bp in length
* 48043 48142: gap of 100 bp
* 48143 48808: contig of 666 bp in length
* 48809 48908: gap of 100 bp
* 48909 49576: contig of 668 bp in length
* 49577 49676: gap of 100 bp
* 49677 50333: contig of 657 bp in length
* 50334 50433: gap of 100 bp
* 50434 51111: contig of 678 bp in length
* 51112 51211: gap of 100 bp
* 51212 51870: contig of 659 bp in length.

```

## Location/Qualifiers

Query Match 0.8%; Score 22; DB 2; Length 51870;

Best Local Similarity 100.0%; Pred. NO. 8.4;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2164 GGAATTCAGGCCATCATGTG 2195

Db 4897 GGAATTCAGGCCATCATGTG 4876

## RESULT 31

## AC129491

## LOCUS

AC129491 Mus musculus clone RP23-66G14, LOW-PASS SEQUENCE SAMPLING.

AC129491.1 GI:22004288

HTG; HTGS PHASE0.

Mus musculus (house mouse)

Mus musculus

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 52855)

Birren, B., Nusbaum, C. and Lander, E.

Mus musculus chromosome, clone RP23-66G14

Unpublished

2 (bases 1 to 52855)

Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,

Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, S.,

Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,

Cook, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J. S., Dodge, S.,

Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J.,

Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B.,

Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A.,

Karas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K.,

Liu, G., MacLean, C., MacDonald, P., Major, J., Matthews, C.,

McCarthy, M., Meldrim, J., Menus, L., Mihova, T., Mienga, V.,

Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H.,

O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,

Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P.,

Roman, J., Roy, A., Schauer, S., Schupbach, R., Seaman, S., Severy, P.,

Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J.,

Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H.,

Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,

Zembek, L., Zimmer, A. and Zody, M.

## TITLE

## JOURNAL

## COMMENT

Direct Submission  
 Submitted (30-JUL-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/SN/RepeatMasker.html>  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: MIBR  
 Web site: <http://www-seq.wi.mit.edu>  
 Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
 ----- Project Information  
 Center project name: L4475  
 Center clone name: 66\_G14  
 -----

\* NOTE: This record contains 66 individual  
 \* sequencing reads that have not been assembled into  
 \* contigs. Runs of N are used to separate the reads  
 \* and the order in which they appear is completely  
 \* arbitrary. Low-pass sequence sampling is useful for  
 \* identifying clones that may be gene-rich and allows  
 \* overlap relationships among clones to be deduced.  
 \* However, it should not be assumed that this clone  
 \* will be sequenced to completion. In the event that  
 \* the record is updated, the accession number will  
 \* be preserved.

```

* 1 709: contig of 709 bp in length
* 710 809: gap of 100 bp
* 810 1530: contig of 721 bp in length
* 1531 1630: gap of 100 bp
* 1631 2337: contig of 707 bp in length
* 2338 2437: gap of 100 bp
* 2438 3144: contig of 707 bp in length
* 3145 3244: gap of 100 bp
* 3245 3946: contig of 702 bp in length
* 3947 4046: gap of 100 bp
* 4047 4755: contig of 709 bp in length
* 4756 4855: gap of 100 bp
* 4856 5543: contig of 688 bp in length
* 5544 5643: gap of 100 bp
* 5644 6325: contig of 682 bp in length
* 6326 6425: gap of 100 bp
* 6426 7146: contig of 721 bp in length
* 7147 7246: gap of 100 bp
* 7247 7945: contig of 699 bp in length
* 7946 8046: gap of 100 bp
* 8046 8738: contig of 693 bp in length
* 8739 8838: gap of 100 bp
* 8839 9549: contig of 711 bp in length
* 9550 9649: gap of 100 bp
* 9649 10334: contig of 685 bp in length
* 10335 10434: gap of 100 bp
* 10435 11153: contig of 719 bp in length
* 11154 11253: gap of 100 bp
* 11254 11969: contig of 716 bp in length
* 11970 12069: gap of 100 bp
* 12070 12778: contig of 709 bp in length
* 12779 12878: gap of 100 bp
* 12879 13593: contig of 715 bp in length
* 13594 13693: gap of 100 bp
* 13694 14402: contig of 709 bp in length
* 14403 14502: gap of 100 bp
* 14503 15172: contig of 670 bp in length
* 15173 15272: gap of 100 bp
* 15273 16057: contig of 685 bp in length
* 15958 16058: gap of 100 bp
* 16058 16764: contig of 706 bp in length
* 16764 16863: gap of 100 bp
* 16864 17561: contig of 698 bp in length
* 17562 17661: gap of 100 bp
* 17662 18347: contig of 686 bp in length
* 18348 18447: gap of 100 bp
* 18448 19141: contig of 694 bp in length

```

McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrim, J., Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phukhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Spencer, B., Stange-Thorann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Triggilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

# Direct Submission

Submitted (06-DEC-2001) Whitehead Institute/MIT Center for Genome Research

Research, 320 Charles Street, Cambridge, MA 02141, USA

All repeats were identified using RepeatMasker:

Smith, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WBIR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L19750

Center clone name: L31\_I\_23

-----

- \* NOTE: This record contains 69 individual
- \* sequencing reads that have not been assembled into
- \* contigs. Runs of N are used to separate the reads
- \* and the order in which they appear is completely
- \* arbitrary. Low-pass sequence sampling is useful for
- \* identifying clones that may be gene-rich and allows
- \* overlap relationships among clones to be deduced.
- \* However, it should not be assumed that this clone
- \* will be sequenced to completion. In the event that
- \* the record is updated, the accession number will
- \* be preserved.

```

1
660: contig of 660 bp in length
661: gap of 100 bp
761: 1418: contig of 658 bp in length
1418: gap of 100 bp
1519: 2177: contig of 659 bp in length
2177: gap of 100 bp
2278: 2936: contig of 659 bp in length
2936: gap of 100 bp
3037: 3695: contig of 659 bp in length
3695: gap of 100 bp
3796: 4445: contig of 650 bp in length
4445: gap of 100 bp
4546: 5205: contig of 660 bp in length
5205: gap of 100 bp
5306: 5954: contig of 649 bp in length
5954: gap of 100 bp
6055: 6694: contig of 640 bp in length
6694: gap of 100 bp
6795: 7475: contig of 681 bp in length
7475: gap of 100 bp
7576: 8232: contig of 657 bp in length
8232: gap of 100 bp
8333: 8963: contig of 631 bp in length
8963: gap of 100 bp
9064: 9730: contig of 667 bp in length
9730: gap of 100 bp
9831: 10482: contig of 652 bp in length
10482: gap of 100 bp
10583: 11225: contig of 643 bp in length
11225: gap of 100 bp
11326: 11976: contig of 651 bp in length
11976: gap of 100 bp
12076: 12732: contig of 656 bp in length
12732: gap of 100 bp
12833: 13497: contig of 665 bp in length
13497: gap of 100 bp
13598: 14264: contig of 667 bp in length
14264: gap of 100 bp
14365: 15112: contig of 648 bp in length
15112: gap of 100 bp
15113: 15763: contig of 651 bp in length
15763: gap of 100 bp
15864: 16514: contig of 651 bp in length
16514: gap of 100 bp
16515: 17376: contig of 662 bp in length
17376: gap of 100 bp
17377: 18045: contig of 669 bp in length
18045: gap of 100 bp
18145: 18816: contig of 671 bp in length
18816: gap of 100 bp
18817: 19366: contig of 650 bp in length
19366: gap of 100 bp
19367: 20334: contig of 668 bp in length
20334: gap of 100 bp
20335: 21082: contig of 648 bp in length
21082: gap of 100 bp
21083: 21734: contig of 552 bp in length
21734: gap of 100 bp
21735: 22486: contig of 652 bp in length
22486: gap of 100 bp
21835: 23241: contig of 655 bp in length
23241: gap of 100 bp
23242: 23991: contig of 650 bp in length
23991: gap of 100 bp
23992: 24742: contig of 651 bp in length
24742: gap of 100 bp
24743: 25507: contig of 665 bp in length
25507: gap of 100 bp
25508: 26261: contig of 654 bp in length
26261: gap of 100 bp
25608: 26351: gap of 100 bp
26262: 27013: contig of 652 bp in length
27013: gap of 100 bp
27014: 27772: contig of 659 bp in length
27772: gap of 100 bp
27773: 28540: contig of 668 bp in length
28540: gap of 100 bp
28541: 28640: gap of 100 bp
28641: 29290: contig of 650 bp in length
29291: 29300: contig of 647 bp in length
29301: 30137: gap of 100 bp
30038: 30790: contig of 653 bp in length
30791: 30890: gap of 100 bp
30891: 31531: contig of 641 bp in length
31532: 31631: gap of 100 bp
31632: 32282: contig of 651 bp in length
32283: 33036: contig of 654 bp in length
33037: 33136: gap of 100 bp
33038: 33799: contig of 663 bp in length
33800: 33899: gap of 100 bp
33900: 34556: contig of 657 bp in length
34557: 34656: gap of 100 bp
34657: 35301: contig of 645 bp in length
35302: 35401: gap of 100 bp
35402: 36054: contig of 653 bp in length
36055: 36154: gap of 100 bp
36155: 36818: contig of 664 bp in length
36819: 36918: gap of 100 bp
36919: 37572: contig of 654 bp in length
37573: 38418: contig of 646 bp in length
38419: 39072: contig of 654 bp in length
39073: 39895: contig of 663 bp in length
39896: 40579: contig of 644 bp in length
40580: 41339: contig of 660 bp in length
41340:

```

BASE COUNT	1372 a	1768 c	1883 g	1091 t
	2702	2335	2723	2356
ORIGIN	Query Match	0.8%;	Score 22;	DB 10; Length 6114;
	Best Local Similarity	100.0%;	Pred. No. 9.9;	
Matches	22;	Conservative	0;	Mismatches 0; Indels 0; Gaps 0;
LOCUS	2702	GCMAAGAGGTCCTCCAGGAGA	2723	
	2335	GCMAAGAGGTCCTCCAGGAGA	2356	
DEFINITION	AF049880	7228 bp	DNA	linear
	Mus musculus	TGFb inducible early protein and hypothetical TGFb		
ACCESSION	AF049880			
	AF049880.1	GI:3882951		
KEYWORDS	Mus musculus	(house mouse)		
	Mus musculus			
SOURCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
AUTHORS	Fautsch, M.P., Vrabel, A., Rickard, D., Subramaniam, M., Spelsberg, T.C.			
	and Wieben, E.D.			
TITLE	Characterization of the mouse TGFbeta-inducible early gene (TIEG);			
	conservation of exon and transcriptional regulatory sequences with			
JOURNAL	evidence of additional transcripts			
	Mamm. Genome 9 (10), 838-842 (1998)			
MEDLINE	98427205			
	9745041			
REFERENCE	2 (bases 1 to 7228)			
	Fautsch, M.P., Vrabel, A., Richard, D., Subramaniam, M., Spelsberg, T.C.			
AUTHORS	and Wieben, E.D.			
	Direct Submission			
TITLE	Submitted (23-FEB-1998) Biochem./Mol. Biology, Mayo			
	Clinic/Foundation, 200 1st St. S.W., Rochester, MN 55905, USA			
FEATURES	Location/Qualifiers			
	1..7228			
source	/organism="Mus musculus"			
	/mol_type="genomic DNA"			
mRNA	/db_xref="taxon:10090"			
	/clone_lib="ES Pl"			
exon	join(795..943,3648..3878,3973..4885,5336..7087)			
	/product="TGFb inducible early protein"			
CDS	795..943			
	/note="1a"			
CDS	join(908..943,3648..3878,3973..4885,5336..5595)			
	/function="transcription factor"			
CDS	/note="TIEG"			
	/codon_start=1			
CDS	/product="TGFb inducible early protein"			
	/protein_id="AAC77797.1"			
CDS	/db_xref="GI:3882952"			
	/translation="MLNQREGKXELISEKPREGMHPKDKAQSDFEAVEALMSXSCDW			
CDS	MSMCDMSKHFKKYLENRPVTSDDTSLLPGTDFVPAFLCTFPYFSDEEPSQSNLT			
	SQGSNLTASASTGHKFSFSDAAKPPGATPKPEEKPLAAPPPLPKAQTAVIRHTAD			
CDS	AQENKQSCPVKAAISILNQNSFRRTTHGVNATRNIPCAAVPNRSKPEPSTVSD			
	QDEKAGALVAVFSPSTVICRSQAPSPVQKSVLSSPTVSTGVPLPVCQV			
CDS	PLPANNLSLSTVTPSPQPAVCSPLFMGTQVPEGTVFVVPQVSPRPVVS			
	PSGRITSLAPAGFSPSAARVTPQIDSSRVSHICSHPGCGKTYFKSHLKAHVTH			
CDS	TGKPFSCSKGGERFARDELSESRHTHTGEKFCACPMCDRFRVSRDHLTKGARRHLSAKKLP			
	LSAKKLPNQMEVSKLNDIALPPTPSAQ"			
CDS	join(1723..1803,3648..3878,3973..4885,5336..7087)			
	/product="hypothetical TGFb inducible early protein splice			
variant	variant"			
	/protein_id="AAC77798.1"			
CDS	/db_xref="GI:3882953"			
	/translation="MLQREGKXELISEKPREGMHPKDKAQSDFEAVEALMSXSCDW			
CDS	ASAPSTGHKFSFSDAAKPPGATPKPEEKPLAAPPPLPKAQTAVIRHTADQCNHQ			
	SCPVKAAISILNQNSFRRTTHGVNATRNIPCAAVPNRSKPEPSTVSDGDEKAGA			
CDS	ALYDFAVRSSSTVICRSQAPSPVQKSVLSSPTVSTGVPLPVCQVPLPANNLS			
	LYSTVTPSPQPAVCSPLFMGTQVPEGTVFVVPQVSPRPVVS			
CDS	PIAPAGFSPSAARVTPQIDSSRVSHICSHPGCGKTYFKSHLKAHVTHGKFPFS			
	SNMGKGERFARDELSESRHTHTGEKFCACPMCDRFRVSRDHLTKGARRHLSAKKLP			
CDS	NMQMEVSKLNDIALPPTPSAQ"			
	1798			
CDS	/note="alternate position for initiation methionine"			
	3648..3878			
CDS	/number=2			
	3973..4885			
CDS	/number=3			
	5336..7087			
CDS	/number=4			
	1611 a	1903 c	1941 g	1773 t
CDS	BASE COUNT			
	ORIGIN			
CDS	Query Match	0.8%;	Score 22;	DB 10; Length 7228;
	Best Local Similarity	100.0%;	Pred. No. 9.7;	
CDS	Matches	22;	Conservative	0; Mismatches 0; Indels 0; Gaps 0;
CDS	QY	1507	GGAGACACCGGCGGAGATGGTG	1528
	DB	179	GGAGACACCGGCGGAGATGGTG	158
CDS	RESULT 30			
	AC104220/c			
CDS	LOCUS	AC104220	51870 bp	DNA linear
	DEFINITION	Mus musculus clone RP23-131123, LOW-PASS SEQUENCE SAMPLING.		
CDS	ACCESSION	AC104220		
	VERSION	AC104220.1	GI:17396386	
CDS	KEYWORDS	HTG; HTGS PHASE0.		
	SOURCE	Mus musculus (house mouse)		
CDS	ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
		Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
CDS	REFERENCE	1 (bases 1 to 51870)		
	AUTHORS	Birren, B., Linton, L., Nusbaum, C. and Lander, E.		
CDS	TITLE	Mus musculus, clone RP23-131123		
	JOURNAL	Unpublished		
CDS	REFERENCE	2 (bases 1 to 51870)		
	AUTHORS	Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,		
CDS		Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B.,		
		Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,		
CDS		Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,		
		Cooke, P., DeArrellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,		
CDS		Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,		
		Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,		
CDS		Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,		
		Jones, C., Kamat, A., Karatas, A., Kells, C., Larocque, K.,		
CDS		Lamazares, R., Landers, T., Lehotzky, J., Levine, R., Liu, G.,		
		MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,		

## RESULT 27

AF272662 5551 bp mRNA linear ROD 11-SEP-2000  
 Rattus norvegicus alpha 1 type V collagen mRNA, complete cds.  
 AF272662  
 DEFINITION  
 Rattus norvegicus alpha 1 type V collagen mRNA, complete cds.  
 AF272662  
 ACCESSION  
 AF272662.1 GI:8568095  
 VERSION  
 1  
 KEYWORDS  
 Rattus norvegicus (Norway rat)  
 SOURCE  
 Rattus norvegicus  
 ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus

## REFERENCE

1 (bases 1 to 5551)  
 Chrennsov,M.A., Rothblum,K., Tyler,W.A., Stahl,R.C. and Carey,D.J.  
 Schwann cells synthesize type V collagen that contains a novel  
 alpha 4 chain. Molecular cloning, biochemical characterization, and  
 high affinity heparin binding of alpha 4(V) collagen

## JOURNAL

J. Biol. Chem. 275 (36), 28208-28215 (2000)

## MEDLINE

20428740

## PUBMED

10852920

## AUTHORS

Chrennsov,M.A., Rothblum,K.N., Tyler,W.A., Stahl,R.C. and  
 Carey,D.J.  
 Direct Submission  
 Submitted (26-MAY-2000) Weis Center for Research, Penn State  
 College of Medicine, 100 N. Academy Ave., Danville, PA 17822, USA

## FEATURES

source

Location/Qualifiers

1..5551  
 /organism="Rattus norvegicus"  
 /mol\_type="mRNA"  
 /strain="Sprague Dawley"  
 /db\_xref="taxon:10116"  
 /cell\_type="Schwann"  
 6..5528  
 /note="collagen alpha chain"  
 /codon\_start=1  
 /product="alpha 1 type V collagen"  
 /protein\_id="AF76433.1"  
 /db\_xref="GI:8568096"

## CDS

/translation="MDVHTRWKAAPRGAPLLSSPLLLLLLLLWAPPRAAQTDLLE  
 MLDHNLPSGVTTTGFCASTRSSRGPDVAVRVSKDAQLSMPTKQLYPDSFDEFSI  
 LTTVKAAGSQAFVSVYNEQGIQQLGLGRSPVLYEDHTGKPGPEYPLDFEINL  
 SDGKWHRIASVYKNNVTLIDCKKIIFLNRGDHPIDVNGIIMFGSRILDDIIFE  
 GDIQLLVSDHRAADYCEHVSPTCDTAVPTPOSQDPNDEYVEGETYVVEYP  
 VYEDPDCKEAPASOKPVEARETTEVPEEQTPRSEAPTVPETSDTAGKEDNSGIG  
 DDYVFDIYTTSPYEDGEGVENPQPTNPSGAEPIFTSVTSNSNPASPE  
 EGKDDLGGEFTETIKNLDENYDPIYFDDSDSNVSPSEIGPMANQDTIYEGIGGP  
 RGEKQKGPALIEPQMLIEGPPGEPAGLPFPPTTGTQMDGPGERGPPRGL  
 PGADGLPFPPTMLPFRFGGSDAGSKGPMVSAQSAQAILQALRALRGAPAGP  
 GLTRGPMPGPGSGGLKGEPMGPGQGVGPPGPTGKPRGRAGSDCARMPG  
 QTPKGDGDFDGLAGLPGKHRGDPGSPGRLGDDGERGDDGVGPRGLGPRGP  
 RGLLPGKPGPPGPGVGTMDQPGPKKNVSPQEPGPPGQCGNPGAGLGPQGA  
 GPPEKGLPGKLPNMGADGPPGHPKGEPPGEGQGGPPGPGPIGYPPRGVKG  
 ADGIRGLTKGEBKSDGFPFGKDMGKNGDGEIGPPGPRGEDGEPGKGRPNQD  
 PGPLPTGKGLGVLPFGPPGPRQPKGSIGPPFPKGANGKGRGTGPKGPRGQR  
 GPTGPRGPRGICITCKPGPKNSGGDPAGPGERGNGPGQPTGFPQPKGPPPG  
 KDLPGKRGORGTGFGKSPGPPGPPGVGPGQGTGTGTPGMRGHPGPPGPGEG  
 PNAAGKEGTGDPGAGLPKQDGPGLRFPFGRDGLPGPVGALGLKSGEGPPGPPGA  
 GSPGEPAGAAAGPFGIPGPPGPPGPPGAGGVPGEPPGQPGAGRDGLQGVGLPG  
 PGPVPPGDESDGIBGEPGKSGDKGEGQPGPTGQPTGQPSGADGPPG  
 RGQGLTGKGLGPPGPPGVLGGLPGPPEKGTGTDVGMGPPGPPGPPGPPG  
 GARGADGPPGPGGGINPAGVAGKGFAGGEPGLPGEGLPGKGEREKEAGTGS  
 AAGCPGKPGDGGKSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG  
 PGTPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG  
 GPDLGLGIPGVGQGLGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG  
 PPEQEGKAGRLGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG  
 TPKGEGAGHLPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG  
 BEIFGSLNLSKLEIQMKRPLQNPARTCKOLQCHPFDGVEVDPNOCSDRS  
 KFYVCKTAGGTCVFPDKKSECARITSWPKENPMSWSEFEKGLLSVYDAEGNPVG  
 VVOMTEFLRLSAAQNIYNVQSVANODATGSDVDAIRFLGNSDEMSTNNPPI  
 RALVGGCATGKQTKVLEIDTPKVEQVFLVDIMFTDFGSEASQKGFGEVPACFLG"

BASE COUNT 1214 a 1645 c 1773 g 919 t

Query Match 0.8%; Score 22; DB 10; Length 5551;  
 Best Local Similarity 100.0%; Pred. No. 9.9;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2702 GCAAGAGAGCTCTCCAGGAGA 2723

Db 2308 GCAAGAGAGCTCTCCAGGAGA 2329

## RESULT 28

CRUPA1V  
 LOCUS CRUPA1V 6114 bp mRNA linear ROD 27-APR-1993  
 DEFINITION Chinese hamster pro-alpha-1 (V) collagen mRNA, complete cds.  
 ACCESSION M76730

VERSION M76730.1 GI:191150

KEYWORDS alpha-1 type V collagen.

SOURCE Cricetulus longicaudatus (long-tailed hamster)

ORGANISM Cricetulus longicaudatus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;

Cricetulus.

1 (bases 1 to 6114)

Greenspan,D.S., Cheng,W. and Hoffman,G.G.

The pro-alpha 1(V) collagen chain. Complete primary structure,

distribution of expression, and comparison with the pro-alpha 1(XI)

collagen chain

J. Biol. Chem. 266 (36), 24727-24733 (1991)

92105142

172213

Original source text: Cricetulus longicaudatus male adult cDNA to

mRNA.

Location/Qualifiers

1..6114  
 /organism="Cricetulus longicaudatus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10030"  
 /sex="male"  
 /dev\_stage="adult"  
 1..6114  
 /gene="pro-alpha-1 (V) collagen"  
 33..5555  
 /gene="pro-alpha-1 (V) collagen"  
 /codon\_start=1  
 /product="pro-alpha-1 type V collagen"  
 /protein\_id="AAA37002.1"  
 /db\_xref="GI:191151"

gene 0

CDS

/translations="MDVHTRWDRLPVGAAPVPLLLLLLLLWAPPQRAAQTDLLE  
 MLDHNLPSGVTTTGFCASTRSSRGPDVAVRVSKDAQLSMPTKQLYPDSFDEFSI  
 LTTVKAAGSQAFVSVYNEQGIQQLGLGRSPVLYEDHTGKPGPEYPLDFEINL  
 SDGKWHRIASVYKNNVTLIDCKKVAFTLNSDHPIDVNGIIMFGSRILDDIIFE  
 GDIQLLVSDHRAADYCEHVSPTCDTAVPTPOSQDPNDEYVEGETYVVEYP  
 YIEDPDCKEAPATKPVARETTEVPEEQTPRSEAPTVPETSDTAGKEDNPGG  
 DDYVFDIYTTAPYEDGEGVENPQPTNPSGAEVPTSTITSNTSNFAPSPG  
 EDKDLGGEFTETIKNLDENYDPIYFDDSDSNVSPSEIGPMANQDTIYEGIGGP  
 RGEKQKGPALIEPQMLIEGPPGEPAGLPFPPTTGTQMDGPGERGPPRGL  
 PGADGLPFPPTMLPFRFGGSDAGSKGPMVSAQSAQAILQALRALRGAPAGP  
 GLTRGPMPGPGSGGLKGEPMGPGQGVGPPGPTGKPRGRAGSDCARMPG  
 QTPKGDGDFDGLAGLPGKHRGDPGSPGRLGDDGERGDDGVGPRGLGPRGP  
 RGLLPGKPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG  
 GPPEKGLPGKLPNMGADGPPGHPKGEPPGEGQGGPPGPPGPIGYPPRGVKG  
 ADGIRGLTKGEBKSDGFPFGKDMGKNGDGEIGPPGPRGEDGEPGKGRPNQD  
 PGPLPTGKGLGVLPFGPPGPPRQPKGSIGPPFPKGANGKGRGTGPKGPRGQR  
 GPTGPRGPRGICITCKPGPKNSGGDPAGPGERGNGPGQPTGFPQPKGPPPG  
 KDLPGKRGORGTGFGKSPGPPGPPGVGPGQGTGTGTPGMRGHPGPPGPGEG  
 PNAAGKEGTGDPGAGLPKQDGPGLRFPFGRDGLPGPVGALGLKSGEGPPGPPGA  
 GSPGEPAGAAAGPFGIPGPPGPPGPPGAGGVPGEPPGQPGAGRDGLQGVGLPG  
 PGPVPPGDESDGIBGEPGKSGDKGEGQPGPTGQPTGQPSGADGPPG  
 RGQGLTGKGLGPPGPPGVLGGLPGPPEKGTGTDVGMGPPGPPGPPGPPG  
 GARGADGPPGPGGGINPAGVAGKGFAGGEPGLPGEGLPGKGEREKEAGTGS  
 AAGCPGKPGDGGKSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG  
 PGTPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG  
 GPDLGLGIPGVGQGLGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG  
 PPEQEGKAGRLGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG  
 TPKGEGAGHLPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG  
 BEIFGSLNLSKLEIQMKRPLQNPARTCKOLQCHPFDGVEVDPNOCSDRS  
 KFYVCKTAGGTCVFPDKKSECARITSWPKENPMSWSEFEKGLLSVYDAEGNPVG  
 VVOMTEFLRLSAAQNIYNVQSVANODATGSDVDAIRFLGNSDEMSTNNPPI  
 RALVGGCATGKQTKVLEIDTPKVEQVFLVDIMFTDFGSEASQKGFGEVPACFLG"





9

Paro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Larocque,K., Lamazares,R., Landers,T., Lehocsky,J., Levine,R., Lindblad-Toh,K., Liu,G., MacLean,C., MacDonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrum,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Riback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

# TITLE

Submitted (06-MAY-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

4 (bases 1 to 188326)

## REFERENCE AUTHORS

Birren,B., Lincon,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,I., Boguslavsky,L., Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B., Choepel,Y., Collangelo,M., Collins,S., Collymore,A., Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Fero,S., Ferreira,P., FitzGerald,M., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Lakocque,K., Lamazares,R., Landers,T., Lehocsky,J., Levine,R., Lindblad-Toh,K., Liu,G., MacLean,C., MacDonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrum,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Riback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

## TITLE

Submitted (31-MAY-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

On May 31, 2002 this sequence version replaced gi:20455631.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RX/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L22138

Center clone name: Y13\_M15

-----

## FEATURES

source

Location/Qualifiers  
1..188326  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="8"  
/map="8"  
/clone="RP11-713M15"  
/clone\_lib="RP11-11 Human Male BAC"  
complement(138..370)  
/rpt\_family="MIR"

repeat\_region

repeat\_region

repeat\_region

/rpt\_family="L1MC4"  
1784..2026  
/rpt\_family="L1MC4"  
2092..2458  
/rpt\_family="L1MC4"  
2480..2532  
/rpt\_family="MERSA"  
complement(3120..3522)  
/rpt\_family="MLT1A1"  
3599..3683  
/rpt\_family="L2"  
3684..3923  
/rpt\_family="MIR"  
4408..4557  
/rpt\_family="MLT1K"  
complement(5098..5184)  
/rpt\_family="MIR3"  
complement(5394..5647)  
/rpt\_family="L2"  
7198..7700  
/rpt\_family="L2"  
7974..8128  
/rpt\_family="MIR3"  
8287..8466  
/rpt\_family="L1PA15"  
complement(8467..8774)  
/rpt\_family="ALUSX"  
8775..8797  
/rpt\_family="L1PA15"  
9069..9274  
/rpt\_family="MIR"  
11640..11666  
/rpt\_family="CGG)n"  
11846..11938  
/rpt\_family="GC-rich"  
12223..12246  
/rpt\_family="GC-rich"  
complement(14375..14433)  
/rpt\_family="MERSA"  
14788..14879  
/rpt\_family="MIR"  
complement(15115..15269)  
/rpt\_family="MIR"  
15270..15571  
/rpt\_family="ALUSg"  
complement(15572..15658)  
/rpt\_family="MIR"  
complement(16044..16174)  
/rpt\_family="MER91C"  
16693..16845  
/rpt\_family="MIR3"  
complement(16848..1746)  
/rpt\_family="MER20B"  
complement(17166..17328)  
/rpt\_family="MER20B"  
complement(17422..17618)  
/rpt\_family="MIR"  
complement(17619..17720)  
/rpt\_family="MIR"  
complement(17749..18274)  
/rpt\_family="MLT1D"  
19131..19174  
/rpt\_family="L2"  
19218..19322  
/rpt\_family="GA-rich"  
20495..20710  
/rpt\_family="MIR"  
21664..21765  
/rpt\_family="L1MED"  
complement(21766..21862)  
/rpt\_family="MER104"  
22027..22086  
/rpt\_family="AT-rich"

```

repeat_region      6930..7275
/rpt_family="MLT1A1"
repeat_region      complement(7405..7475)
/rpt_family="MIR"
repeat_region      7482..7550
/rpt_family="TTCC)n"
repeat_region      complement(8801..8907)
/rpt_family="MIR"
repeat_region      complement(9382..9417)
/rpt_family="MIR"
repeat_region      9433..9519
/rpt_family="LIMCb"
repeat_region      9704..9756
/rpt_family="T-rich"
repeat_region      complement(9814..9976)
/rpt_family="MIR"
repeat_region      10289..10609
/rpt_family="AluSx"
repeat_region      10692..10819
/rpt_family="MER58C"
repeat_region      12811..13083
/rpt_family="AluJo"
repeat_region      13574..13598
/rpt_family="AT-rich"
repeat_region      13952..14081
/rpt_family="LTR33"
repeat_region      14920..15040
/rpt_family="MER58A"
repeat_region      15041..15474
/rpt_family="Tigger2a"
repeat_region      15475..15555
/rpt_family="MER58A"
repeat_region      16115..16262
/rpt_family="LIPAL3"
repeat_region      complement(16263..16553)
/rpt_family="AluSx"
repeat_region      16553..16811
/rpt_family="LIPAL3"
repeat_region      17748..17989
/rpt_family="MIR"
repeat_region      18007..18036
/rpt_family="AT-rich"
repeat_region      19080..19136
/rpt_family="TTCC)n"
repeat_region      19190..19265
/rpt_family="L2"
repeat_region      20639..20817
/rpt_family="MIR"
repeat_region      21219..21259
/rpt_family="TA)n"
repeat_region      21378..21958
/rpt_family="LIPAL5"
repeat_region      22511..22552
/rpt_family="L2"
repeat_region      complement(22606..22655)
/rpt_family="L2"
repeat_region      23202..23475
/rpt_family="AluJb"
repeat_region      23498..23656
/rpt_family="MIR"
repeat_region      24106..24135
/rpt_family="TG)n"
repeat_region      24136..24180
/rpt_family="CA)n"
repeat_region      24589..24643
/rpt_family="TAA)n"
repeat_region      complement(24710..24970)
/rpt_family="MIR"
repeat_region      24993..25068
/rpt_family="L3"
repeat_region      25110..25285
/rpt_family="L2"
repeat_region      complement(26026..26154)

```

```

/rpt_family="MIR"
26450..26538
/rpt_family="L3"
complement(27038..27100)
/rpt_family="MIR"
complement(27912..28052)
/rpt_family="FLAM-C"
complement(28946..29209)
/rpt_family="MER58B"
29210..29844
/rpt_family="LINCx"
29845..29885
/rpt_family="AT-rich"
30076..30343
/rpt_family="LLXCa"
31695..31801
repeat_region

```

Query Match 0.88, Score 23, DB 9, Length 166093;  
 Best Local Similarity 100.0%; Pred. No. 2.2;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1450 CCAGGTGTTCCAGGATCCGAGG 1472
      |||||
DB 75280 CCAGGTGTTCCAGGATCCGAGG 75258

```

## RESULT 24

```

AC104958      188326 bp      DNA      linear      PRI 31-MAY-2002
LOCUS      Homo sapiens chromosome 9, clone RP11-713M15, complete sequence.
AC104958
VERSION      AC104958.6      GI:21281562
KEYWORDS      HTG.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Hominidae; HOMO.
REFERENCE      1 (bases 1 to 188326)
AUTHORS      Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N.,
TITLE      Homo sapiens chromosome 9, clone RP11-713M15
JOURNAL      Unpublished
REFERENCE      2 (bases 1 to 188326)
AUTHORS      Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazare,B.,
Chapel,Y., Colangelo,M., Collins,S., Collamore,A., Cook,A.,
Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Fato,S.,
Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,M., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
Lamazares,R., Landers,T., Lehotzky,J., Levine,R., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrum,J.,
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phukhang,P., Pierre,N., Pollara,V.,
Rydmond,F., Retta,R., Rieback,M., Riley,K., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupbach,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Triggilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zairour,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (22-DEC-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
1 (bases 1 to 188326)
REFERENCE
AUTHORS

```

```

repeat_region 29327..29456
/vidences=not_experimental
/rpt_family="WTR16C"
29510..29865
/vidences=not_experimental
/rpt_family="HAL1"
repeat_region complement(29895..30257)
/vidences=not_experimental
/rpt_family="MER119"
repeat_region 30514..30739
/vidences=not_experimental
/rpt_family="HAL1"
repeat_region 30742..31096
/vidences=not_experimental
/rpt_family="HAL1"
repeat_region 31126..31174
/vidences=not_experimental
/rpt_family="AT_rich"
repeat_region 31572..31631
/vidences=not_experimental
/rpt_family="AT_rich"
repeat_region 31796..31892
/vidences=not_experimental

Query Match 0.8%; Score 23; DB 9; Length 151880;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1450 CCAGTGTTCAGGATCCAGG 1472
|||||
DB 36464 CCAGTGTTCAGGATCCAGG 36442

RESULT 23
AC106013/C
LOCUS Homo sapiens chromosome 8, clone RP11-1109L18, complete sequence.
DEFINITION AC106013
VERSION AC106013.2 GI:18543063
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 166093)
Birken, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Boguski, L., Boukhgalter, B.,
Brown, A., Camarata, J., Campoliano, A., Chang, J., Chazaro, B.,
Choepe, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
Cooke, P., DeArellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,
Ferrel, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K.,
Lamazzari, R., Lander, E., Landers, T., Lehotzky, J., Levine, R., Liu, G.,
MacLean, C., Macdonald, P., McKernan, K., McPheeters, R., Meldrim, J.,
Meneus, L., Minova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Topham, K., Travers, M., Travis, N., Trigilio, J., Tsefaye, S.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Submitted (11-JAN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
-----
Center: Whitehead Institute/MIT Center for Genome Research
Center code: WISR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
-----
Center project name: L22151
Center clone name: 1109_L18
-----
FEATURES
Source
repeat_region 1..166093
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/map="8"
/clone="RP11-1109L18"
/clone_lib="RP11-11 Human Male BAC"
repeat_region 1..103
/rpt_family="AluSg/x"
complement(550..747)
/rpt_family="MIR3"
repeat_region 1559..1838
/rpt_family="AluSx"
2812..2969
/rpt_family="L2"
2913..3032
/rpt_family="MIR"
3123..3423
/rpt_family="AluSx"
complement(3449..3646)
/rpt_family="MIR"
4330..4742
/rpt_family="MLT1B"
complement(4899..5264)
/rpt_family="THE1C"
5877..5929
/rpt_family="CT-rich"
6015..6149
/rpt_family="MIR3"
complement(6127..6218)
/rpt_family="MIR"

```

```

Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 166093)
Birken, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Boguski, L., Boukhgalter, B.,
Brown, A., Camarata, J., Campoliano, A., Chang, J., Chazaro, B.,
Choepe, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
Cooke, P., DeArellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,
Ferrel, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kells, C., LaRocque, K., Lamazzari, R.,
Landers, T., Lehotzky, J., Levine, R., Liu, G., MacLean, C.,
Macdonald, P., McKernan, K., McPheeters, R., Meldrim, J., Meneus, L.,
Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Topham, K., Travers, M., Travis, N., Trigilio, J., Tsefaye, S.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (02-MAR-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA.
On Feb 6, 2002 this sequence version replaced gi:18129519.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
-----
Center: Whitehead Institute/MIT Center for Genome Research
Center code: WISR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
-----
Center project name: L22151
Center clone name: 1109_L18
-----
FEATURES
Source
repeat_region 1..166093
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/map="8"
/clone="RP11-1109L18"
/clone_lib="RP11-11 Human Male BAC"
repeat_region 1..103
/rpt_family="AluSg/x"
complement(550..747)
/rpt_family="MIR3"
repeat_region 1559..1838
/rpt_family="AluSx"
2812..2969
/rpt_family="L2"
2913..3032
/rpt_family="MIR"
3123..3423
/rpt_family="AluSx"
complement(3449..3646)
/rpt_family="MIR"
4330..4742
/rpt_family="MLT1B"
complement(4899..5264)
/rpt_family="THE1C"
5877..5929
/rpt_family="CT-rich"
6015..6149
/rpt_family="MIR3"
complement(6127..6218)
/rpt_family="MIR"

```

TITLE  
JOURNAL

```

/cell_type="pre-pro-B cell"
/clone_lib="Keio BAC library"
708..999
/evidence=not_experimental
/rpt_family="AluSq"
complement(2018..2340)
/evidence=not_experimental
/rpt_family="AluSq"
complement(2418..2723)
/evidence=not_experimental
/rpt_family="AluSq"
3205..3242
/evidence=not_experimental
/rpt_family="At_rich"
3413..3550
/evidence=not_experimental
/rpt_family="AluJo"
complement(3669..3835)
/evidence=not_experimental
/rpt_family="L2"
complement(3860..4092)
/evidence=not_experimental
/rpt_family="MIR"
complement(4300..4502)
/evidence=not_experimental
/rpt_family="L2"
complement(4566..4680)
/evidence=not_experimental
/rpt_family="L2"
complement(4707..4805)
/evidence=not_experimental
/rpt_family="L2"
complement(4806..5130)
/evidence=not_experimental
/rpt_family="AluSq"
complement(5131..5172)
/evidence=not_experimental
/rpt_family="L2"
5783..5954
/evidence=not_experimental
/rpt_family="GGAA)n"
complement(6062..6171)
/evidence=not_experimental
/rpt_family="L1P45"
6172..6325
/evidence=not_experimental
/rpt_family="L1P2"
6376..6524
/evidence=not_experimental
/rpt_family="MIR3"
complement(6712..8802)
/evidence=not_experimental
/rpt_family="L2"
9129..9182
/evidence=not_experimental
/rpt_family="(TTTA)n"
9400..9713
/evidence=not_experimental
/rpt_family="AluSc"
complement(9958..10357)
/evidence=not_experimental
/rpt_family="L2"
11068..11090
/evidence=not_experimental
/rpt_family="At_rich"
11338..11530
/evidence=not_experimental
/rpt_family="HALL"
complement(11702..11817)
/evidence=not_experimental
/rpt_family="MIR3"
complement(12101..12524)
/evidence=not_experimental

```

```

repeat_region
12595..12627
/evidence=not_experimental
/rpt_family="At_rich"
12761..12800
/evidence=not_experimental
/rpt_family="(CATA)n"
13340..13360
/evidence=not_experimental
/rpt_family="At_rich"
13493..13513
/evidence=not_experimental
/rpt_family="At_rich"
13765..13913
/evidence=not_experimental
/rpt_family="t2"
complement(13914..14159)
/evidence=not_experimental
/rpt_family="MER2"
complement(14160..14470)
/evidence=not_experimental
/rpt_family="AluSq"
complement(14471..14564)
/evidence=not_experimental
/rpt_family="MER2"
14565..14617
/evidence=not_experimental
/rpt_family="L2"
14860..16488
/evidence=not_experimental
/rpt_family="L1P4"
15463..20818
/evidence=not_experimental
/rpt_family="L1P43"
20819..20839
/evidence=not_experimental
/rpt_family="(TAAAA)n"
20840..22278
/evidence=not_experimental
/rpt_family="L1P43"
22279..22327
/evidence=not_experimental
/rpt_family="(CA)n"
22328..22674
/evidence=not_experimental
/rpt_family="L1P43"
complement(25183..25311)
/evidence=not_experimental
/rpt_family="MER53"
25424..25457
/evidence=not_experimental
/rpt_family="(A)n"
complement(25797..26148)
/evidence=not_experimental
/rpt_family="MLTIC"
26149..26522
/evidence=not_experimental
/rpt_family="MSTA"
complement(26523..26625)
/evidence=not_experimental
/rpt_family="MLTIC"
27718..27750
/evidence=not_experimental
/rpt_family="(?)n"
complement(27893..28065)
/evidence=not_experimental
/rpt_family="MIR"
28344..28496
/evidence=not_experimental
/rpt_family="MER5B"
28515..28731
/evidence=not_experimental
/rpt_family="MER20B"

```

```

HLTFs"
12881..15512
/notes="Similar to putative retrotransposon,
Ty3-gypsy-like"
complement(12881..14462)
/notes="Similar to putative retrotransposon,
Ty3-gypsy-like"
complement(15519..15766)
/notes="Similar to putative retrotransposon,
Ty3-gypsy-like"
15828..19063
/genes="OSJNB0079E01.3"
/notes="Putative retroelement"
join(15828..15923,16528..16774,17501..17614,17716..17782,
17809..18342,18655..18878,19002..19063)
/genes="OSJNB0079E01.3"
/notes="Similar to sp|P04323|POL3_DROME Retrovirus-related
pol polyprotein from transposon 17.6 . E-value=7e-32"
/codon_start=1
/product="Putative retroelement"
/protein_id="AAW74387.1"
/db_xref="GI:21672025"
/translation="MTCSQDGNSPQLKNNFEYFVYVHPKRWELLILHPDYDF
VEAELEDEGNLMAISQAAVGTESATSFRLRGWVQTKVLMFVNSRSGHSIDSL
AHRPFSSILLQRLRYGHQIRLMGEERHTAFQSHGVYRVMSFQLTDMOSHLSH
LYKVFQKXWSEVYWLHGWHLVVKGSLITKIEVVNKKVPTSVKLRFRFGLAGY
LYKVFQKGLISKPLTHLLRKDVSVFSPADQAFQALKALVTVPLVALLDYSKFT
IATDGDAGVILCOQGHPIVNAKALGVKSRGLSTYKEYELAILLDVDRHSYLOH
AEFILLDSHLSMLSEQGIYHWMKGTFTVHNVIEVHEPLMGRCQAVGAIAFWL
EGPEWLRSRSGQGTCTILQHLFALHPSKRLPQARQLGVGPHHAFFPTSPRARVR"
17448..17631
/notes="Similar to Oryza sativa gypsy-type retrotransposon
RIRB135 DNA, partial sequence"
complement(17907..18054)
/notes="Similar to putative retrotransposon,
Ty3-gypsy-like"
complement(18113..31694)
/notes="Similar to putative retrotransposon,
Ty3-gypsy-like"
23972..28537
/genes="OSJNB0079E01.4"
/notes="Putative retroelement"
join(23972..24771,24790..27749,27846..28222,28198..28537)
/genes="OSJNB0079E01.4"
/notes="Similar to sp|Q05654|RDPO_SCHPO Retrotransposable
element 112 155 kDa protein . E-value=e-121; Contains pfam
domain: rvt Reverse transcriptase (RNA-dependent) D
e-value=7.9e-54; Contains pfam domain: rve Integrase core
domain e-value=9.1e-19; Contains pfam domain: zf-CCHC Zinc
knuckle e-value=3.1e-06"
/codon_start=1
/product="Putative retroelement"
/protein_id="AAW74388.1"
/db_xref="GI:21672026"
/translations="MYTTRNGSRATGESNGEERTEGTVHPNSNGGPPPLPENPLA
QVWAHQTMAMMMQOQHQHQMCHAEQQHQFQPPFQSKLPEFLVRPPT
FSSTTPMEANDWHAIEKKXNLLQCNDEKRYAFATHQLQGFASAMNDHMAPEPPT
EVTVAEFCSFKKQAPDGVAKQKREFRALHQGNRTVTEYLFNFRLARYAPEVRT
DAEKQKFLAGLDELDTNLISGDYAFKAIROEDQNRQNRKRAPOGSHORP
RFAPGQCGPTTMIHQYRFPNPSNPFQAGSGSNHGGQPNRGAAPPMAPQSGP
SAQAKETGAKGSCFCNGCELGHFADKCPKPRACRFQIARVNHASAEQAAPVW
LGTFPNYSIPATLIPDSGATHSFISKFFVGMHGLKEELTEPMVHTPGSSVSFES
PSULIEIQRSFRLANILLESKDLVLLQMDLTKFGVILCANRTVTLTNEQETVW
YKSPDSFGQVSLNQIAELIPVTEVNNLRKQEDIPVCSIFPEFFEDLTTPPKKEI
BFRIDLAPGATVYKRPYMAANAEAVKQVDEQKQYIRLTSFPMGAPVIFVEK
DKTRKMCVIRALNEVTIRKNTPLRIDDLFDLQKGVKFDKILRSYGHLRIRED
IPKTAFTTRYGLCECTVMSFGLTNAPAFPMILMNKVFMEFLDKFVVFVFDLIIYSKS
EEBHQRLVLEKLEKHOLYAKFSCDFMLTEVFLGHVITAQGVAVQPSNVESYTK
WTPPKTSQILRSKGLAGYRRFIENFSRIAPMTQLLKDEKFTACDCKSFEELX
KXLSAPVILLPQMDQVQYCDASPHGLGCVLMQEGRYAVASRQLRPHENYPFH
LELAANVHALKIWRHLYGNRECVYTDHKSLEYIFIQDLNLRRLWLELIKOTWSI
HYHFGKAVADALSRKSYCTALCTIEGMCEELEHNLVIEHGFVALEARPPL
VDQVRAQVNDNSIABELKKNMVRVGKARDFHEDEHGTIWLGERLCVPDCKELKDILTE

```

```

AHOCTOYSHPGSTKMYQDLKSKFWMVSMRREIAEFVVLCDVQCVKAEKHORPAGLLQOP
LOIQEMKEEIGMDIFGLPPTSSCHDSIYVWVDLTKVAHFIPVHTTGTGKEAELEY
LSRIMCUGVPKKIKVSDRGSOFTSKFNQKLAELGCTRLNESTAYHISOIDBFSSNNYSQ
ASLQMAPFALYGRKCRTPFLPDQTDERQDLFGTEVLAEBEAKVKRIIRERLRIASQK
SVADNRRELTIFAGDYVYLVRVTLRGEVAYQLELPSNMIGIHDFVHVSOLKCLRP
EQADSEHIDIQEDLTVEYKPVRIIDTSERRTRNKVTRFCVQMSHSEETATWERD
ELKAAHPHLFASSESSESRGDSV"
31695..33497
/genes="OSJNB0079E01.5"
/notes="Hydrothermal protein"
join(31695..31947,31920..31995,32085..32971,33132..33497)
/genes="OSJNB0079E01.5"
/codon_start=1
/product="Hypothetical protein"
/protein_id="AAW74389.1"
/db_xref="GI:21672027"
/translation="MARWFYHTIPFEAGSESAKALRCRRRAIAFNRKPKIADVGAMEA
RPALLCKSQVTVDOGEVDGLPKLVDPGNTVLTACANAARPMIDDDVYVYSOL
LTRYAVYRANRVYNGKLPWMPANPKVDDDDGVSFKTRGOVKLAPRRVPTSDSDA
DDEDDVKERCNEEGEBEKEAAVVEKAADSVENCVDTEGYTTPSPGHNERTGVSN
SPLPRKDLGKALVAFSVGKTANGPVKISKKGLVDVAVFSDNDESSDSTLSP
AGHSLDLSTAPILSLDAGGVGSAAGALASADRIVMAAARVFGSPURQVW9PLVNA
KGGAAVETSAKSYSLAAPHFAPGDFETRADLVFVKGVLAHLRAKDDDEIGYKSL
MEALANTKEAKENKRLQTELEKGEARAEVDRLNAKLEKEKAHATVL7DYVNLTEP
KMEALLEGSKAKASAVESORLLREMAKATESTKTACQCPY"
complement(35071..35117)
/notes="Simple repeat (A)n"
complement(35200..39069)
/genes="OSJNB0079E01.6"
/notes="Putative retroelement"
Query Watch
Best Local Similarity 100.0%; Score 23; DB 8; Length 132870;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1647 AAAAGGTGAAAAGCGGGAATGCTG 1669
|||||:|||||:|||||:|||||:
Db 49858 AAAAGGTGAAAAGCGGGAATGCTG 49836

RESULT 22
AP005716/c
AP005716
LOCUS Homo sapiens genomic DNA, chromosome 8q23, clone: KB44559,
DEFINITION complete sequence.
ACCESSION AP005716
VERSION AP005716.2 GI:28411858
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 Shimizu.N. and Asakawa.S.
AUTHORS Shimizu.N. and Asakawa.S.
TITLE Homo sapiens DNA chromosome 8 SEQUENCE
JOURNAL Published Only in Database (2002)
REFERENCE 2 (bases 1 to 151880)
AUTHORS Shimizu.N. and Asakawa.S.
TITLE Direct Submission
JOURNAL Submitted (07-SEP-2002) Nobuyoshi Shimizu, Keio University, School
of Medicine, Molecular Biology, 35 Shinanomachi, Shinjuku-ku, Tokyo
160-8582, Japan (E-mail:nshimizugmb.med.keio.ac.jp,
Tel:81-3-3351-2370, Fax:81-3-3351-2370)
COMMENT On Feb 18, 2003 this sequence version replaced gi:22779366.
FEATURES
Location/Qualifiers
source
1..151880
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="9"
/map="8q23"
/clone="KB144559"
/cell_line="FLEB 14 - 14"

```

\* 41180 contig of 723 bp in length  
 \* 42002 gap of 100 bp  
 \* 42713 contig of 711 bp in length  
 \* 42813 gap of 100 bp  
 \* 43515 contig of 702 bp in length  
 \* 43616 gap of 100 bp  
 \* 44314 contig of 699 bp in length  
 \* 44414 gap of 100 bp  
 \* 45122 contig of 708 bp in length  
 \* 45123 gap of 100 bp  
 \* 45223 contig of 721 bp in length  
 \* 45944 gap of 100 bp  
 \* 46044 contig of 681 bp in length  
 \* 46725 gap of 100 bp  
 \* 46824 gap of 100 bp  
 \* 47552 contig of 728 bp in length  
 \* 47553 gap of 100 bp  
 \* 47653 contig of 728 bp in length  
 \* 48381 gap of 100 bp  
 \* 48481 contig of 718 bp in length  
 \* 49199 gap of 100 bp  
 \* 49299 contig of 722 bp in length  
 \* 50021 gap of 100 bp  
 \* 50836 contig of 716 bp in length  
 \* 50837 gap of 100 bp  
 \* 50937 contig of 719 bp in length  
 \* 51655 gap of 100 bp  
 \* 51756 gap of 100 bp  
 \* 52470 contig of 714 bp in length  
 \* 52569 gap of 100 bp  
 \* 53270 contig of 701 bp in length  
 \* 53271 gap of 100 bp  
 \* 54076 contig of 706 bp in length  
 \* 54176 gap of 100 bp  
 \* 54177 contig of 681 bp in length  
 \* 54857 gap of 100 bp  
 \* 54858 gap of 100 bp  
 \* 55643 contig of 686 bp in length  
 \* 55644 gap of 100 bp  
 \* 55744 contig of 717 bp in length

Query Match 0.84; Score 23; DB 2; Length 67884;

Best Local Similarity 100.0%; Pred. No. 2.3;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1450 CCAGGTGTCAGGATCTCCAGG 1472  
 Db 25961 CCAGGTGTCAGGATCTCCAGG 25939

RESULT 21  
 AC119149/c

LOCUS AC119149 132870 bp DNA linear PLN 14-JUL-2002  
 DEFINITION Oryza sativa (japonica cultivar-group) chromosome 10 clone  
 AC119149

ACCESSION AC119149.2 GI:21306633  
 VERSION 2  
 KEYWORDS HTG.

SOURCE Oryza sativa (japonica cultivar-group)

ORGANISM Oryza sativa (japonica cultivar-group)  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE 1 (bases 1 to 132870)  
 AUTHORS Ferraro, K., Kuit, K., Nascimento, L., Zucavern, T., Ballia, V.,  
 Bell, M., Baker, J., Miller, B., Katzenberger, F., Muller, S., King, L.,  
 Sullivan, P., Yang, C., Dike, S., O'Shaughnessy, A., Palmer, L. and  
 Dedhia, N.

TITLE Genomic sequence for Oryza sativa, Nipponbare strain, clone  
 OSJNB0079E01, from chromosome 10, complete sequence

JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 132870)

AUTHORS McCombie, W.R.

TITLE Direct Submission

JOURNAL Submitted (25-APR-2002) Lita Annenberg Hazen Genome Center, Cold

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

Spring Harbor Laboratories, 1, Bungtown Road, Cold Spring Harbor,  
 NY 11724, USA  
 3 (bases 1 to 132870)  
 McCombie, W.R.  
 Direct Submission  
 Submitted (01-JUN-2002) Lita Annenberg Hazen Genome Center, Cold  
 Spring Harbor Laboratories, 1, Bungtown Road, Cold Spring Harbor,  
 NY 11724, USA

REFERENCE  
 AUTHORS

4 (bases 1 to 132870)  
 Palmer, L.E., de la Bastide, M., Spiegel, L., Prestor, R., Ferraro, K.,  
 Kuit, K., Nascimento, L., Zucavern, T., Ballia, V., Bell, M., Baker, J.,  
 Miller, B., Katzenberger, F., Muller, S., King, L., Sullivan, P.,  
 Yang, C., Dike, S., O'Shaughnessy, A., Dedhia, N. and McCombie, W.R.  
 Direct Submission

TITLE  
 JOURNAL

Submitted (02-JUL-2002) Lita Annenberg Hazen Genome Center, Cold  
 Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY  
 11724, USA

REMARK Oryza sativa (japonica cultivar-group) chromosome 10 clone  
 OSJNB0079E01, complete sequence.

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

5 (bases 1 to 132870)  
 McCombie, W.R.  
 Direct Submission  
 Submitted (14-JUL-2002) Genome Research Center, Cold Spring Harbor  
 Laboratory, 500 Sunnyside Blvd, Woodbury, NY 11797, USA  
 On Jun 1, 2002 this sequence version replaced gi:20103540.

COMMENT

This sequence was finished as follows unless otherwise noted: all  
 regions were either double-stranded or sequenced with an alternate  
 chemistry or covered by high quality data (i.e., phred quality >=  
 30); an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by at least  
 one plasmid subclone or more than one M13 subclone; and the  
 assembly was confirmed by restriction digest. The nucleotide  
 sequence of this BAC clone was generated by combining fragments and  
 sequence of this BAC clone was generated by combining fragments and  
 Cold Spring Harbor Laboratory Genome Center sequencing data.  
 Clone OSJNB0079E01 overlaps clone OSJNB0053D03 (AC021892) from  
 base 1 to base 35377. The overlap is from base 16377 to base  
 199113 on OSJNB0053D03. Clone OSJNB0079E01 also overlaps clone  
 OSJNB0036B06 (AC051634) from base 121446 to base 132870. The  
 overlap is from base 1 to base 11425 on OSJNB0036B06.

FEATURES

source

1..132870  
 /organism="Oryza sativa (japonica cultivar-group)"

/mol\_type="genomic DNA"

/cultivar="Nipponbare"

/db\_xref="taxon:39947"

/chromosome="10"

/clone="OSJNB0079E01"

/clone\_lib="EccRI"

/complement(2679..2700)

/note="Low complexity AT-rich"

2738..2914

/note="Simple repeat (CCGG)n"

/complement(6746..6946)

/note="Similar to unknown retrotransposon"

complement(6942..7610)

/note="Similar to unknown retrotransposon"

complement(9396..9395)

/note="Similar to putative transposon, CACTA, En/Spm-like"

complement(10340..10384)

/note="Low complexity GC-rich"

12175..12204

/note="Simple repeat (GA)n"

12544..12867

/gene="OSJNB0079E01.2"

/note="Hypothetical protein"

12544..12867

/gene="OSJNB0079E01.2"

/codon\_start=1

/product="Hypothetical protein"

/protein\_id="AA074386.1"

/db\_xref="GI:21672024"

/translation="MAAKVLDQKQKDLKKYEDSDRKESEKRESESEKSRIF

NALKVVESHFTVVKRVDELQGVLEQKVEKLQIVVQQAAREPFSAPTPTGAI

S

0



```

TITLE
JOURNAL
COMMENT
Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,A., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
McPheeters,R., Meldrum,J., Meneus,L., Mihova,T., Mkena,V.,
Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
Roy,A., Santos,R., Schauer,S., Schuback,R., Seaman,S., Severy,P.,
Sougnuez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Travers,M., Travis,N., Trigglio,J., Vassiliev,H., Viel,R., Vo,A.,
Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zaitoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (03-MAR-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L11533
Center clone name: 567_N_24
-----
* NOTE: This record contains 84 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
* 1 684: contig of 684 bp in length
*
* 685 784: gap of 100 bp
*
* 785 1507: contig of 723 bp in length
*
* 1508 1607: gap of 100 bp
*
* 1608 2320: contig of 713 bp in length
*
* 2321 2420: gap of 100 bp
*
* 2421 3131: contig of 711 bp in length
*
* 3132 3231: gap of 100 bp
*
* 3232 3907: contig of 676 bp in length
*
* 3908 4007: gap of 100 bp
*
* 4008 4711: contig of 704 bp in length
*
* 4712 4811: gap of 100 bp
*
* 4812 5512: contig of 701 bp in length
*
* 5513 5612: gap of 100 bp
*
* 5613 6316: contig of 704 bp in length
*
* 6317 6416: gap of 100 bp
*
* 6417 7135: contig of 719 bp in length
*
* 7136 7235: gap of 100 bp
*
* 7236 7954: contig of 719 bp in length
*
* 7955 8054: gap of 100 bp
*
* 8055 8776: contig of 722 bp in length
*
* 8777 9593: contig of 717 bp in length
*
* 9594 9694: gap of 100 bp
*
* 9694 10411: contig of 718 bp in length
*
* 10412 10511: gap of 100 bp
*
* 10512 11224: contig of 713 bp in length
*
* 11225 11324: gap of 100 bp
*
* 11325 12028: contig of 704 bp in length
*
*
* 12029 12128: gap of 100 bp
*
* 12129 12806: contig of 678 bp in length
*
* 12807 12906: gap of 100 bp
*
* 12907 13610: contig of 704 bp in length
*
* 13611 13710: gap of 100 bp
*
* 13711 14436: contig of 726 bp in length
*
* 14437 14536: gap of 100 bp
*
* 14537 15263: contig of 727 bp in length
*
* 15264 15363: gap of 100 bp
*
* 15364 16080: contig of 717 bp in length
*
* 16081 16180: gap of 100 bp
*
* 16181 16989: contig of 709 bp in length
*
* 16990 16989: gap of 100 bp
*
* 16991 17708: contig of 719 bp in length
*
* 17709 17808: gap of 100 bp
*
* 17810 18531: contig of 723 bp in length
*
* 18532 18631: gap of 100 bp
*
* 18632 19323: contig of 692 bp in length
*
* 19324 19423: gap of 100 bp
*
* 19424 20130: contig of 707 bp in length
*
* 20131 20230: gap of 100 bp
*
* 20231 20925: contig of 695 bp in length
*
* 20926 21025: gap of 100 bp
*
* 21026 21728: contig of 703 bp in length
*
* 21729 21828: gap of 100 bp
*
* 21829 22521: contig of 693 bp in length
*
* 22522 23221: gap of 100 bp
*
* 23223 23314: contig of 693 bp in length
*
* 23315 23414: gap of 100 bp
*
* 23415 24146: contig of 732 bp in length
*
* 24147 24246: gap of 100 bp
*
* 24247 24961: contig of 715 bp in length
*
* 24962 25061: gap of 100 bp
*
* 25062 25771: contig of 710 bp in length
*
* 25772 25871: gap of 100 bp
*
* 25872 26570: contig of 699 bp in length
*
* 26571 26670: gap of 100 bp
*
* 26671 27388: contig of 718 bp in length
*
* 27389 27488: gap of 100 bp
*
* 27489 28192: contig of 704 bp in length
*
* 28193 28292: gap of 100 bp
*
* 28293 28999: contig of 707 bp in length
*
* 28999 29099: gap of 100 bp
*
* 29099 29807: contig of 708 bp in length
*
* 29808 29907: gap of 100 bp
*
* 29908 30600: contig of 693 bp in length
*
* 30601 30700: gap of 100 bp
*
* 30701 31403: contig of 703 bp in length
*
* 31404 31503: gap of 100 bp
*
* 31504 32219: contig of 716 bp in length
*
* 32220 32319: gap of 100 bp
*
* 32321 33034: contig of 715 bp in length
*
* 33035 33134: gap of 100 bp
*
* 33135 33855: contig of 721 bp in length
*
* 33856 33955: gap of 100 bp
*
* 33956 34658: contig of 703 bp in length
*
* 34659 34759: gap of 100 bp
*
* 34759 35461: contig of 703 bp in length
*
* 35462 35561: gap of 100 bp
*
* 35562 36266: contig of 704 bp in length
*
* 36267 36365: gap of 100 bp
*
* 36366 37082: contig of 717 bp in length
*
* 37083 37182: gap of 100 bp
*
* 37183 37868: contig of 686 bp in length
*
* 37869 37968: gap of 100 bp
*
* 37969 38673: contig of 705 bp in length
*
* 38674 38773: gap of 100 bp
*
* 38774 39462: contig of 689 bp in length
*
* 39463 39562: gap of 100 bp
*
* 39563 40254: contig of 692 bp in length
*
* 40255 40354: gap of 100 bp
*
* 40355 41079: contig of 725 bp in length
*
* 41080 41179: gap of 100 bp

```

```

repeat_region 8435. .8737
/note="AluX repeat: matches 1. .298 of consensus"
repeat_region 8738. .9144
/note="L1M2 repeat: matches 4193. .4614 of consensus"
repeat_region 9339. .11100
/note="MER66-internal repeat: matches 4895. .6676 of consensus"
repeat_region 11101. .11496
/note="MER66B repeat: matches 1. .392 of consensus"
repeat_region 11510. .12032
/note="L1M2 repeat: matches 4604. .5114 of consensus"
repeat_region 12033. .12338
/note="AluX repeat: matches 1. .306 of consensus"
repeat_region 12339. .13410
/note="L1M2 repeat: matches 5114. .6307 of consensus"
repeat_region 13633. .13852
/note="AluYb repeat: matches 87. .307 of consensus"
repeat_region 13896. .15355
/note="TIGER2 repeat: matches 1266. .2718 of consensus"
repeat_region 15356. .15657
/note="AluX repeat: matches 1. .301 of consensus"
repeat_region 15658. .16040
/note="TIGER2 repeat: matches 863. .1266 of consensus"
repeat_region 16041. .16350
/note="AluYb repeat: matches 1. .306 of consensus"
repeat_region 16351. .16417
/note="TIGER2 repeat: matches 798. .863 of consensus"
repeat_region 16445. .16538
/note="FLAM C repeat: matches 24. .117 of consensus"
repeat_region 16579. .16625
/note="TIGER2 repeat: matches 1. .47 of consensus"
repeat_region 16749. .17174
/note="MER65A repeat: matches 1. .445 of consensus"
repeat_region 17696. .17996
/note="AluX repeat: matches 1. .306 of consensus"
repeat_region 18497. .18587
/note="FLAM A repeat: matches 43. .130 of consensus"
repeat_region 18596. .18897
/note="AluX repeat: matches 1. .302 of consensus"
repeat_region 19209. .19517
/note="AluX repeat: matches 1. .312 of consensus"
repeat_region 20539. .20659
/note="MIR repeat: matches 98. .228 of consensus"
repeat_region 20912. .21038
/note="AluYb repeat: matches 1. .126 of consensus"
repeat_region 21039. .21330
/note="AluX repeat: matches 1. .291 of consensus"
repeat_region 21331. .21467
/note="AluYb repeat: matches 126. .266 of consensus"
repeat_region 21473. .21756
/note="AluYb repeat: matches 14. .295 of consensus"
repeat_region 22440. .22524
/note="AluX repeat: matches 1. .286 of consensus"
repeat_region 23046. .23585
/note="L2 repeat: matches 2234. .2745 of consensus"
repeat_region 23679. .24214
/note="MER41A repeat: matches 1. .538 of consensus"
repeat_region 24274. .24440
/note="MIR repeat: matches 17. .191 of consensus"
repeat_region 24622. .24927
/note="AluX repeat: matches 1. .305 of consensus"
repeat_region 26536. .26843
/note="AluYb repeat: matches 1. .303 of consensus"
repeat_region 27436. .27726
/note="AluYb repeat: matches 13. .304 of consensus"
repeat_region 29235. .29339
/note="L2 repeat: matches 2596. .2705 of consensus"
repeat_region 29588. .29879
/note="AluX repeat: matches 1. .293 of consensus"
repeat_region 30927. .31173
/note="LTR23 repeat: matches 204. .437 of consensus"
repeat_region 31638. .31785
/note="MER5A repeat: matches 87. .189 of consensus"
repeat_region 31786. .31914
/note="FLAM C repeat: matches 1. .131 of consensus"
repeat_region 31915. .31989
/note="MER5A repeat: matches 2. .87 of consensus"
repeat_region 32211. .32511
/note="AluX repeat: matches 1. .301 of consensus"
repeat_region 33285. .33395
/note="MER81 repeat: matches 3. .114 of consensus"
repeat_region 33429. .33566
/note="FLAM C repeat: matches 1. .133 of consensus"
repeat_region 33567. .33606
/note="20 copies 2 mer tg 87% conserved"
repeat_region 34452. .34750
/note="AluX repeat: matches 1. .299 of consensus"
repeat_region 34798. .35089
/note="AluYb repeat: matches 1. .292 of consensus"
repeat_region 35591. .35905
/note="AluX repeat: matches 1. .312 of consensus"
repeat_region 35917. .35978
/note="L2 repeat: matches 2622. .2690 of consensus"
repeat_region 36699. .37024
/note="AluX repeat: matches 1. .311 of consensus"
repeat_region 37215. .37349
/note="L1M5A repeat: matches 6146. .6295 of consensus"
repeat_region 37451. .37586
/note="FLAM C repeat: matches 1. .133 of consensus"
repeat_region 37718. .37742
/note="MER5B repeat: matches 73. .96 of consensus"
repeat_region 37743. .38041
/note="AluX repeat: matches 1. .301 of consensus"
repeat_region 38042. .38272
/note="MER5B repeat: matches 96. .336 of consensus"
repeat_region 38504. .38786
/note="AluX repeat: matches 6. .306 of consensus"
repeat_region 39348. .40556
/note="CpG island"
misc_feature 40556. .43073
/evidence=not experimental
complement(42531. .43073)
/note="match: STS: Em:G36503"
repeat_region 45639. .45763
/note="L2 repeat: matches 2617. .2750 of consensus"
repeat_region 47289. .47314
/note="13 copies 2 mer ca 100% conserved"
repeat_region 47413. .47506
Query Match 0.8; Score 24; DB 9; Length 128960;
Best Local Similarity 100.0%; Pred. No. 0.62;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Cy 2191 GGAGAGAGGTGAAGAGGAGAA 2214
Db 90972 GGAGAGAGGTGAAGAGGAGAA 90995
|||||
RESULT 20
AC090561/c
LOCUS AC090561 67884 bp DNA linear HTG 03-MAR-2001
DEFINITION Homo sapiens clone RP11-567N24, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC090561
VERSION 1
KEYWORDS HTG; HTGS PHASE0.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 67884)
Birten,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens, clone RP11-567N24
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 67884)
Birten,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
Barra,N., Bastien,V., Boguslavsky,L., Boukhgalter,B., Brown,A.,
Camarata,J., Campopiano,A., Chcepel,Y., Colangelo,G.M., Collins,S.,

```

```

repeat_region /note="AluSq repeat: matches 1. .306 of consensus"
32417. .32806
repeat_region /note="L1M4 repeat: matches 4603. .5002 of consensus"
32818. .33181
repeat_region /note="L1M1 repeat: matches -1388. .-1025 of consensus"
33182. .33493
repeat_region /note="AluSp repeat: matches 1. .313 of consensus"
33494. .34874
repeat_region /note="L1M1 repeat: matches -1025. .355 of consensus"
34876. .34981
repeat_region /note="L1P repeat: matches 3001. .3103 of consensus"
34986. .35806
repeat_region /note="L1P2 repeat: matches 5275. .6144 of consensus"
35825. .41028
repeat_region /note="L1M1 repeat: matches 347. .5615 of consensus"
41029. .41334
repeat_region /note="AluSq repeat: matches 6. .310 of consensus"
41335. .41728
repeat_region /note="L1M1 repeat: matches 5615. .6005 of consensus"
41727. .41963
repeat_region /note="L1M2 repeat: matches 5825. .6063 of consensus"
41976. .42095
repeat_region /note="60 copies 2 mer at 85% conserved"
42097. .42330
repeat_region /note="L1M1 repeat: matches 6081. .6322 of consensus"
42547. .42818
repeat_region /note="L1M7 repeat: matches 5778. .6033 of consensus"
42819. .43022
repeat_region /note="AluY repeat: matches 102. .303 of consensus"
43023. .43157
repeat_region /note="L1M7 repeat: matches 6033. .6171 of consensus"
43283. .43474
repeat_region /note="L1R repeat: matches 21. .247 of consensus"
43140. .45454
repeat_region /note="AluY repeat: matches 1. .311 of consensus"
45525. .45828
repeat_region /note="AluSq repeat: matches 1. .307 of consensus"
47590. .47773
repeat_region /note="L1M1 repeat: matches 1309. .1491 of consensus"
47984. .48257
repeat_region /note="L1M1 repeat: matches 630. .931 of consensus"
49161. .49464
repeat_region /note="AluSq repeat: matches 1. .305 of consensus"
50177. .50264
repeat_region /note="22 copies 4 mer agag 70% conserved"
50186. .50263
repeat_region /note="39 copies 2 mer ga 74% conserved"
50465. .50490
repeat_region /note="13 copies 2 mer gc 96% conserved"
50465. .50490
BASE COUNT 17751 a 9364 c 8310 g 15187 t
ORIGIN
Query Match 2.0%; Score 58; DB 9; Length 50612;
Best Local Similarity 100.0%; Pred. No. 1.3e-19;
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1586 AGGTCGAATGGTGCACCAAGGAGAGCAAAAGGATCACCCTGGATTATGGCAAAAGGG 1643
|||||
Db 46132 AGGTCGAATGGTGCACCAAGGAGAGCAAAAGGATCACCCTGGATTATGGCAAAAGGG 46075
|||||

RESULT 19
HSJ919F19 128960 bp DNA linear PRI 12-JUN-2001
LOCUS Human DNA sequence from clone RP5-919F19 on chromosome 6q16.3-22.1,
DEFINITION complete sequence.
ACCESSION AL109947
VERSION AL109947.19 GI:13897493
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

REFERENCE Johnson, C  
Direct Submission  
Submitted (11-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
requests: clonerequest@sanger.ac.uk  
On Apr 30, 2001 this sequence version replaced gi:13751877.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest. The following  
abbreviations are used to associate primary accession numbers given  
in the feature table with their source databases: EM, EMBL; SW, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; information on the WORMPEP  
database can be found at  
http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence  
was generated from part of bacterial clone contigs of human  
chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping  
Group. Further information can be found at  
http://www.sanger.ac.uk/MGP/Chr6  
RP5-919F19 is from the library RPC1-5 constructed by the group of  
Peter de Jong. For further details see  
http://www.chori.org/bacpac/home.htm  
VECTOR: pCVPAC2  
This sequence is the entire insert of clone RP5-919F19. The true  
left end of clone RP5-1036D8 is at 86977 in this sequence. The true  
right end of clone RP11-425D10 is at 56374 in this sequence.

FEATURES  
location/Qualifiers  
1. .128960  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/map="q16.3-22.1"  
/clone="RP5-919F19"  
/clone\_1b="RPC1-5"  
repeat\_region 16. .330  
/note="AluY repeat: matches 1. .298 of consensus"  
repeat\_region 878. .1177  
/note="AluY repeat: matches 1. .301 of consensus"  
repeat\_region 1409. .1704  
/note="AluX repeat: matches 1. .296 of consensus"  
repeat\_region 2306. .2424  
/note="FLAM\_A repeat: matches 1. .123 of consensus"  
repeat\_region 3935. .4225  
/note="AluSg repeat: matches 1. .299 of consensus"  
repeat\_region 4512. .4698  
/note="L1MCS repeat: matches 7347. .7517 of consensus"  
repeat\_region 5185. .5506  
/note="AluJo repeat: matches 1. .308 of consensus"  
repeat\_region 5558. .6086  
/note="L1MCS repeat: matches 6834. .7364 of consensus"  
repeat\_region 6137. .6330  
/note="L1MC/D repeat: matches 5276. .5467 of consensus"  
repeat\_region 6393. .6686  
/note="AluY repeat: matches 1. .301 of consensus"  
repeat\_region 6998. .7112  
/note="FLAM\_A repeat: matches 1. .115 of consensus"  
repeat\_region 7186. .7588  
/note="HAL1 repeat: matches 505. .914 of consensus"  
repeat\_region 7998. .8355  
/note="L1MA2 repeat: matches 3613. .3976 of consensus"  
repeat\_region 8372. .8434  
/note="L1MA2 repeat: matches 4133. .4193 of consensus"



```

TITLE
JOURNAL
COMMENT
Sequence tag and encoded human protein
Patent: JP 2001269182-A 29842 02-OCT-2001;
GENSET
OS Homo sapiens (human)
PN JP 2001269182-A/29842
DD 02-OCT-2001
PF 24-FEB-2000 JP 2000118773
PR 26-FEB-1999 US 60/122487
PI JEAN BAPTISTE DUMAS MILNE EDWARDS, EIMERIC DUCLAIR, JEAN YVES
PI JORDAN
PC C12N15/09, C07K14/435, C07K16/18, C12N1/15, C12N1/19, C12N1/21, PC
C12N5/10
PC C12P21/02, C12P21/08, C12Q1/68, G06F17/30, C12N15/00, C12N5/00, PC
G06F15/40
CC

FEATURES
    source
        1..246
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /db_xref="taxon:9606"
BASE COUNT
    70 a 63 c 71 g 40 t 2 others
ORIGIN
    Query Match
    Best Local Similarity 100.0%; Pred. No. 1.5e-59; Length 246;
    Matches 131; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2722 GACCAGGTTCCCTGGCAAGATGGAGACCATGGAATCCAAAGGCAACCA 2751
DB 112 GACCAGGTTCCCTGGCAAGATGGAGACCATGGAATCCAAAGGCAACCA 171
QY 2782 GCCCCCCAGGATCTGGACCATCATCTATGTTTAGTGAATGGCAGAGAGATCCG 2841
DB 172 GCCCCCCAGGATCTGGACCATCATCTATGTTTAGTGAATGGCAGAGAGATCCG 231
QY 2842 TTCAGAAAAGG 2852
DB 232 TTCAGAAAAGG 242

RESULT 16
AX721212
LOCUS
DEFINITION
    Sequence 172 from Patent WO0220754.
ACCESSION
    AX721212
VERSION
    AX721212.1 GI:30422048
KEYWORDS
    Homo sapiens (human)
ORGANISM
    Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
    1
    Lincoln, S.E., Altus, C.M., Dufour, G.E., Chalup, M.S., Hillman, J.L.,
    Jones, A., Yu, J.Y., Wright, R.J., Gietzen, D., Liu, Toomy, P., Yap, P.,
    Dahl, C.R., Moniyama, M., Bradley, D., Rohatgi, S., Harris, B.,
    Roseberry, A.M., Gerstein, E.H., Peralta, C.H., David, M., Panzer, S.,
    Flores, V., Dafo, A., Marwaha, R., Chen, A., Chang, S.C. and Inman, R.R.
    Molecules for diagnostics and therapeutics
    Patent: WO 0220754-A 172 14-MAR-2002;
    Incyte Genomics, Inc. (US)
FEATURES
    source
        1..1101
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /note="Incyte ID No: LI:1184621.4:2000SEP08"
BASE COUNT
    334 a 191 c 203 g 373 t
ORIGIN
    Query Match
    Best Local Similarity 100.0%; Pred. No. 3.2e-22; Length 1101;
    Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2195 AGAGAGGTGAAAGGGAGAACTCGGTCCGAGTCCCATTTGGATCAAAAGGAGATCTG 2254
DB 140 AGAGAGGTGAAAGGGAGAACTCGGTCCGAGTCCCATTTGGATCAAAAGGAGATCTG 199
QY 2255 GGG 2257
DB 200 GGG 202

RESULT 17
AL450431
LOCUS
DEFINITION
    Human DNA sequence from clone RP11-323114 on chromosome 6, complete
    sequence.
ACCESSION
    AL450431 AL39091
VERSION
    AL450431.1 GI:11493128
KEYWORDS
    HTG.
SOURCE
    Homo sapiens (human)
ORGANISM
    Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
    1 (Bases 1 to 25483)
    Williams, S.
    Direct Submission
    Submitted (28-MAY-2002) Wellcome Trust Sanger Institute, Hinxton,
    Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
    hinxton@sanger.ac.uk
    During sequence assembly data is compared from overlapping clones.
    Where differences are found these are annotated as variations
    together with a note of the overlapping clone name. Note that the
    variation annotation may not be found in the sequence submission
    corresponding to the overlapping clone, as we submit sequences with
    only a small overlap as described above.
    The following abbreviations are used to associate primary accession
    numbers given in the feature table with their source databases:
    Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information
    on the WORMPEP database can be found at
    http://www.sanger.ac.uk/projects/C.elegans/wormpep
    This sequence
    was generated from part of bacterial clone contigs of human
    chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
    Group. Further information can be found at
    http://www.sanger.ac.uk/MCP/chr6
    This sequence was finished as follows unless otherwise noted: all
    regions were either double-stranded or sequenced with an alternate
    chemistry or covered by high quality data (i.e., phred quality >=
    30); an attempt was made to resolve all sequencing problems, such
    as compressions and repeats; all regions were covered by at least
    one plasmid subclone or more than one M13 subclone; and the
    assembly was confirmed by restriction digest. RP11-323114 is from
    the library RP11-11.2 constructed by the group of Pieter de Jong.
    For further details see
    VECTO: aBACE3.6.
    Location/Qualifiers
        1..25483
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosome="6"
            /clone="RP11-323114"
            /clone_lib="RP11-11.2"
BASE COUNT
    7053 a 4365 c 4610 g 9455 t
ORIGIN
    Query Match
    Best Local Similarity 100.0%; Pred. No. 1.4e-19;
    Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1586 AGGGTGAATGGGTCCCAAGGAGAGCAAGGATCCTCGATTTATGCGCAAAAGGG 1643
DB 2625 AGGGTGAATGGGTCCCAAGGAGAGCAAGGATCCTCGATTTATGCGCAAAAGGG 2682
```

Db 205 ATCTTAGAGCTCTAAATTTGTTGGCCAGAAAATTTGAATAGTGAAGAAGTGGCTT 264  
 QY 181 GTCAATATCACAAATACTTTGACATAGGCGCGAAGTTTATTCAGTTGAGTGGTTCAA 240  
 Db 265 GTCATATACAAAATACTTTGACATAGGCGCGAAGTTTATTCAGTTGAGTGGTTCNA 324  
 QY 241 TATAGTGACTACCTGCTGGAGATTCTCTCGGAAGCTATGATTCAGGAGAATTTG 300  
 Db 325 TATAGTGACTACCTGCTGGAGATTCTCTCGGAAGCTATGATTCAGGAGAATTTG 384  
 QY 301 ACGCACAGTGGATCATCTACTTACTTAGGAGGAACACACAGACAGGGAAG 354  
 Db 385 ACGCAGCAGTGGGAATCACTACTTACTTAGGAGGAACACACAGACAGGGAAG 438

RESULT 13  
 AF438327.3  
 WPCOMMENT

Sequence split into 4 fragments LOCUS AF438327 Accession AF438327

Fragment Name	Begin	End
AF438327.0	1	110000
AF438327.1	100001	210000
AF438327.2	200001	310000
AF438327.3	300001	381696

Continuation (4 of 4) of AF438327 from base 300001 (AF438327 Homo sapiens alpha 1 type X

Query Match 7.1k; Score 204; DB 9; Length 81696;  
 Best Local Similarity 100.0k; Pred. No. 6.8e-100;  
 Matches 204; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2397 AGCCAGCTACCACTCTTACTTCAGAGTGAAGAATTAGAAATTTGATCTTGCCTGTC 2456  
 Db 47106 AGCCAGCTACCACTCTTACTTCAGAGTGAAGAATTAGAAATTTGATCTTGCCTGTC 47165  
 QY 2457 CCACATGGCTCCCGGGTATTCCTGGCCACCTGTCGATAGGCGCCAGAGGTCGCG 2516  
 Db 47166 CCACATGGCTCCCGGGTATTCCTGGCCACCTGTCGATAGGCGCCAGAGGTCGCG 47225  
 QY 2517 AGGATTACTCTGTTTGCAGGAAGAGATGGTGTTCCTGGATTAGTGGGTGTCCTGGAGC 2576  
 Db 47226 AGGATTACTCTGTTTGCAGGAAGAGATGGTGTTCCTGGATTAGTGGGTGTCCTGGAGC 47285  
 QY 2577 TCCAGGTGTCAGAGATTAAAGG 2600  
 Db 47286 TCCAGGTGTCAGAGATTAAAGG 47309

RESULT 14  
 HS682J15/C  
 LOCUS  
 DEFINITION Human DNA sequence from clone RP4-682J15 on chromosome 6p11.2-12.3, complete sequence.  
 ACCESSION AL034452  
 VERSION AL034452.8 GI:4678496  
 KEYWORDS HTG.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 Smith.S.

Direct Submission  
 Submitted (04-MAR-2003) Wellcome Trust Sanger Institute, Hinxton,  
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
 humquery@sanger.ac.uk  
 On Apr 24, 1999 this sequence version replaced gi:4586169.

----- Genome Center  
 Center: Wellcome Trust Sanger Institute  
 Center code: SC  
 Web site: <http://www.sanger.ac.uk>  
 Contact: [humquery@sanger.ac.uk](mailto:humquery@sanger.ac.uk)  
 -----

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one X13 subclone; and the rare assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:

Em., EMBL; Swi, SWISSPROT; Tr., TrEMBL; Wp., WORMPEP; Information

on the WORMPEP database can be found at

[http://www.sanger.ac.uk/projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/projects/C_elegans/wormpep) This sequence

was generated from part of bacterial clone contigs of human

chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping

Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/Chr6>

RP4-682J15 is from the library RPI-4 constructed by the group of

Pieter de Jong. For further details see

<http://www.chori.org/bacpac/home.htm>

VECTOR: pCYPAC2.

FEATURES  
 source  
 1..118431  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="6"  
 /map="p11.2-12.3"  
 /clone="RP4-682J15"  
 /clone\_lib="RPI-4"

BASE COUNT 38100 a 21308 c 21316 g 37707 t  
 ORIGIN

Query Match 7.1k; Score 204; DB 9; Length 118431;  
 Best Local Similarity 100.0k; Pred. No. 6.6e-120;  
 Matches 204; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2397 AGCCAGCTACCACTCTTACTTCAGAGTGAAGAATTAGAAATTTGATCTTGCCTGTC 2456  
 Db 99561 AGCCAGCTACCACTCTTACTTCAGAGTGAAGAATTAGAAATTTGATCTTGCCTGTC 99002

QY 2457 CCACATGGCTCCCGGGTATTCCTGGCCACCTGTCGATAGGCGCCAGAGGTCGCG 2516  
 Db 99501 CCACATGGCTCCCGGGTATTCCTGGCCACCTGTCGATAGGCGCCAGAGGTCGCG 98942

QY 2517 AGGATTACTCTGTTTGCAGGAAGAGATGGTGTTCCTGGATTAGTGGGTGTCCTGGAGC 2576  
 Db 98941 AGGATTACTCTGTTTGCAGGAAGAGATGGTGTTCCTGGATTAGTGGGTGTCCTGGAGC 98882

QY 2577 TCCAGGTGTCAGAGGATTAAAGG 2600  
 Db 98881 TCCAGGTGTCAGAGGATTAAAGG 98858

RESULT 15  
 BD053596  
 LOCUS  
 DEFINITION Sequence tag and encoded human protein.  
 ACCESSION BD053596  
 VERSION BD053596.1 GI:25599202  
 KEYWORDS GP 2001269182-A/29842.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 Edwards.J.B.D.M., Duclair,E. and Jordan,J.Y.

REFERENCE  
 1. (bases 1 to 246)  
 Edwards.J.B.D.M., Duclair,E. and Jordan,J.Y.

```

QY 209 GGCAGAGTTTATTCAAGTTGGAGTGGTTCAATATAGTGACATACCTCTGTGCTGGAGATTC 268
Db 50007 GGCAGAGTTTATTCAAGTTGGAGTGGTTCAATATAGTGACATACCTCTGTGCTGGAGATTC 49948
QY 269 CTCTCGAGAGTATGATTCAGGAGAACATTTCAGCGGAGCAGTGGAAATCCATCTACT 328
Db 49947 CTCTCGAGAGTATGATTCAGGAGAACATTTCAGCGGAGCAGTGGAAATCCATCTACT 49888
QY 329 TAGGAGAAACACAAAGACAGGAGGCCATCCAGTTTCGCGTCGATACCTTTTGGCA 388
Db 49887 TAGGAGAAACACAAAGACAGGAGGCCATCCAGTTTCGCGTCGATACCTTTTGGCA 49828
QY 389 AGTCCTCACCATTCTGACTACATAGCAGTGGTCTTACGATGGCAATCCCAAGATG 448
Db 49827 AGTCCTCACCATTCTGACTACATAGCAGTGGTCTTACGATGGCAATCCCAAGATG 49768
QY 449 ACCTCAAGATGAGCTCAAGCAGCAGAGAGATAGTAAGATAACATTATTTGCTATTGGTG 508
Db 49767 ACCTCAAGATGAGCTCAAGCAGCAGAGAGATAGTAAGATAACATTATTTGCTATTGGTG 49702
QY 509 TTGGTTTCAAGAACAGAGATGCCAGCTTAGAGCTATTCGCCAAGAGCCTTCGCTACTT 568
Db 49707 TTGGTTTCAAGAACAGAGATGCCAGCTTAGAGCTATTCGCCAAGAGCCTTCGCTACTT 49648
QY 569 ATGTGTTTTATGTGGAGAGCTATATTGCAATATCCAAATAGGGAAGTGTAGAGCAGA 628
Db 49647 ATGTGTTTTATGTGGAGAGCTATATTGCAATATCCAAATAGGGAAGTGTAGAGCAGA 49588
QY 629 AACTTTGTGAAG 640
Db 49587 AACTTTGTGAAG 49576

RESULT 11
AX721209 1414 bp mRNA linear PAT 07-MAY-2003
LOCUS
DEFINITION Sequence 169 from Patent WO0220754.
ACCESSION AX721209
VERSION AX721209.1 GI:30422045
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE
AUTHORS Lincoln,S.E., Altus,C.M., Dufour,G.E., Chalup,M.S., Hillman,J.L., Jones,A., Yu,J.Y., Wright,R.J., Gietzen,D., Liu, Toomy,F., Yap,P., Dahl,C.R., Momiyaama,M., Bradley,D., Rohatgi,S., Harris,B., Roseberry,A.M., Gerstin,E.H., Peralta,C.H., David,M., Panzer,S., Flores,V., Daffo,A., Marwaha,R., Chen,A., Chang,S.C. and Inman,R.R.
Molecules for diagnostics and therapeutics
Patent: WO 0220754-A 169 14-MAR-2002;
Incyte Genomics, Inc. (US)
FEATURES
source
1. .1414
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/notes="Incyte ID No: LG:407788.2:2000SEP08"
BASE COUNT 425 a 267 c 295 g 427 t
ORIGIN
Query Match 13.3%; Score 382; DB 6; Length 1414;
Best Local Similarity 100.0%; Pred. No. 1.2e-197;
Matches 382; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2397 AGCCAGCTACAGCTTCTACTTCAGAGTGGAGAAATTAGAAATTGTGATCATGCTCTC 2456
Db 959 AGCCAGCTACAGCTTCTACTTCAGAGTGGAGAAATTAGAAATTGTGATCATGCTCTC 10:8
QY 2457 CCAACATGGCTCCCGGGTATTCCTCGGCCACCTGTCGATAGCCAGAGGTCGCCAG 2516
Db 2457 CCAACATGGCTCCCGGGTATTCCTCGGCCACCTGTCGATAGCCAGAGGTCGCCAG 2516

Db 2019 CCAACATGGCTCCCGGGTATTCCTCGGCCACCTGTCGATAGCCAGAGGTCGCCAG 1079
QY 2517 AGGATTACCTGGTTTGCAGGAAAGAGATGGTTTCTTGGATTAGTGGTGTCTCTCGAAG 2576
Db 1078 AGGATTACCTGGTTTGCAGGAAAGAGATGGTTTCTTGGATTAGTGGTGTCTCTCGAAG 1138
QY 2577 TCCAGGTGTGAGAGATTAAAGGCTTACAGGAGAAATGGGMAAAGGAGCCAGG 2636
Db 1139 TCCAGGTGTGAGAGATTAAAGGCTTACAGGAGAAATGGGMAAAGGAGCCAGG 1198
QY 2637 GTTTCGATATCTGAGAACAAAGGTCTCTCTGTTCCGCCAGGTCCAGAGGCGCTCTCTGG 2696
Db 2199 GTTTCGATATCTGAGAACAAAGGTCTCTCTGTTCCGCCAGGTCCAGAGGCGCTCTCTGG 1258
QY 2697 AATAGCAAGAGAGGTCTCTCAGAGACCCAGGTCTCTCTGGCAAGATGGAGACCATG 2756
Db 1259 AATAGCAAGAGAGGTCTCTCAGAGACCCAGGTCTCTCTGGCAAGATGGAGACCATG 1318
QY 2757 AAAACCTGGAATCCCAAGGGCAA 2778
Db 1319 AAAACCTGGAATCCCAAGGGCAA 1340

RESULT 12
BD058397
LOCUS
DEFINITION Secreted expressed sequence tags (sESTs).
ACCESSION BD058397
VERSION BD058397.1 GI:22604003
KEYWORDS JP 2001519666-A/252.
SOURCE Zea mays
ORGANISM Zea mays
REFERENCE
AUTHORS Jacobs,K., McCoy,J.M., Lavalie,E.R., Racie,L.A., Marberg,D., Treacy,M., Spaulding,V. and Agostino,M.J.
TITLE Secreted expressed sequence tags (sESTs)
JOURNAL Patent: JP 2001519666-A 252 23-OCT-2001;
GENETICS INSTITUTE INC
COMMENT PN JP 2001519666-A/252
BD 23-OCT-2001
PF 10-APR-1998 JP 1998543068
PR 10-APR-1997 US 08/835913
PI KENNETH JACOBS,JOHN M MCCOY,EDWARD R LAVALLIE,LISA A RACIE,PI DAVID MARBERG,
PI MAURICE TREACY,VIRIKI SPAULDING,MICHAEL J AGOSTINO PC
C12N15/12,C12N5/10,C07K14/47,C12Q1/68,A61K38/17 CC Strandedness:
Double:
CC Topology: Linear;
FH Key Location/Qualifiers.
source
1. .447
/organism="Zea mays"
/mol_type="genomic DNA"
/db_xref="taxon:4577"
BASE COUNT 141 a 80 c 101 g 125 t
ORIGIN
Query Match 8.8%; Score 252; DB 6; Length 447;
Best Local Similarity 99.4%; Pred. No. 4.1e-126;
Matches 352; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGGCTCCTACTATATACATTTCTCTGGATGGTTTGGTGGTCTCTCAGATTCCTG 60
Db 85 ATGGCTCCTACTATATACATTTCTCTGGATGGTTTGGTGGTCTCTCAGATTCCTG 144
QY 61 TTAGCTGAAGATGGGGAAGTAAGATCAAGTTGTCTACTGCTCCGACAGATTTAGTTTC 120
Db 145 TTAGCTGAAGATGGGGAAGTAAGATCAAGTTGTCTACTGCTCCGACAGATTTAGTTTC 204
QY 121 ATCTTAGATGGCTCTTATAGTGGTGGCCGCAAAACTTTGAAATAGTGAAGAGTGGCT 180

```

Db 915 AAAAAAGTTTAAAGAAAGATACAGCTTTCCACAAAAAAGATAAAGGATATGAAGTAACA 974

Qy 781 TCAGAAAGTTGATTTATCAGAACTCAAGCAATGTTTCCCAAGAGGCTTCCTCCCATCA 840

Db 975 TCAGAAAGTTGATTTATCAGAACTCAAGCAATGTTTCCCAAGAGGCTTCCTCCCATCA 1034

Qy 841 TATGATTTTGTCTACTCAAGATTTAAAGTCAAGAAATTTGGATTTATGAGAAAT 930

Db 1035 TATGATTTTGTCTACTCAAGATTTAAAGTCAAGAAATTTGGATTTATGAGAAAT 1094

Qy 901 TTAACATTGATGGAAGGCCA 921

Db 1095 TTAACATTGATGGAAGGCCA 1115

RESULT 9

AF438327.2

WPCOMMENT

Sequence split into 4 fragments LOCUS AF438327 Accession AF438327

Fragment Name Begin End

AF438327\_0 100001 110000

AF438327\_1 100001 210000

AF438327\_2 200001 310000

AF438327\_3 300001 381696

Continuation (3 of 4) of AF438327 from base 200001 (AF438327 Homo sapiens alpha 1 type X

Query Match 19.3% Score 552; DB 9; Length 110000;

Best Local Similarity 100.0%; Pred. No. 3e-291;

Matches 552; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 89 GTTGTGCTACTGCTCCGACAGATTTAGTTTTCATCTTAGATGGCTTCCTATAGTGTGGCC 148

Db 27182 GTTGTGCTACTGCTCCGACAGATTTAGTTTTCATCTTAGATGGCTTCCTATAGTGTGGCC 27241

Qy 149 CAGAAACTTTGAAATAGTGAAGAGTGGCTTGTCAATATCAGAAACTTTGACATAG 208

Db 27242 CAGAAACTTTGAAATAGTGAAGAGTGGCTTGTCAATATCAGAAACTTTGACATAG 27301

Qy 209 GGCCTCAAGTTTATCAAGTTGGAGTGGTTCAATATAGTGAATACCTGTGCTGAGATTC 268

Db 27302 GGCCTCAAGTTTATCAAGTTGGAGTGGTTCAATATAGTGAATACCTGTGCTGAGATTC 27361

Qy 269 CTCTCGAAGCTATGATTCAGGAGAACATTTGACGCGACGAGTGAATCCATCTACT 328

Db 27362 CTCTCGAAGCTATGATTCAGGAGAACATTTGACGCGACGAGTGAATCCATCTACT 27421

Qy 329 TAGGAGGAACACAAAGACAGGAGGCCATCCAGTTTTCGCTGCTGATTTACCTTTTGCCA 388

Db 27422 TAGGAGGAACACAAAGACAGGAGGCCATCCAGTTTTCGCTGCTGATTTACCTTTTGCCA 27481

Qy 389 AGTCTCAGCAATTTCTGACTAAGATAGCAGTGGTACTTACGATGGCAATCCGAGATG 448

Db 27482 AGTCTCAGCAATTTCTGACTAAGATAGCAGTGGTACTTACGATGGCAATCCGAGATG 27541

Qy 449 ACCTCAAGGATGAGCTCAAGCAGCAGAGAGATAGTGAATACATTTTGTATTTGCTG 508

Db 27542 ACCTCAAGGATGAGCTCAAGCAGCAGAGAGATAGTGAATACATTTTGTATTTGCTG 27601

Qy 509 TTGTTTCAGAAACAGAGATGCCGAACTTAGAGCTATTGCGACAGGCTTCGCTACTT 568

Db 27602 TTGTTTCAGAAACAGAGATGCCGAACTTAGAGCTATTGCGACAGGCTTCGCTACTT 27661

Qy 569 ATGTGTTTTATGTGGAAGATATATTGCAATATCCAAATTAAGGGAAGTGAAGAGCAGA 628

Db 27662 ATGTGTTTTATGTGGAAGATATATTGCAATATCCAAATTAAGGGAAGTGAAGAGCAGA 27721

Qy 629 AACTTTGTGAAG 640

Db 27722 AACTTTGTGAAG 27733

RESULT 10

HS708F5/c

LOCUS

DEFINITION

Human DNA sequence from clone RP4-708F5 on chromosome 6p11.2-12.3, complete sequence.

ACCESSION

AL031782

VERSION

AL031782.1 GI:4164360

KEYWORDS

HTG.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

REFERENCE

1. (bases 1 to 126501)

Griffiths, C.

Direct Submission

TITLE

Submitted (04-MAR-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk

AUTHORS

Humquerry@sanger.ac.uk

JOURNAL

On Jan 19, 1999 this sequence version replaced gi:3947543.

COMMENT

-----

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: humquerry@sanger.ac.uk

-----

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WormPEP; Information on the WormPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/chr6> RP4-708F5 is from the library RPC1-4 constructed by the group of Peter de Jong. For further details see <http://www.cebri.org/bacpac/home.htm> VECTOR: pCVPAC2.

FEATURES

source

1..126501

Location/Qualifiers

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/chromosome="6"

/map="p11.2-12.3"

/clone="RP4-708F5"

/clone\_lib="RPC1-4"

BASE COUNT 44177 a 23609 c 21902 g 36813 t

ORIGIN

Query Match. 19.3% Score 552; DB 9; Length 126501;

Best Local Similarity 100.0%; Pred. No. 2.9e-291;

Matches 552; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 89 GTTGTGCTACTGCTCCGACAGATTTAGTTTTCATCTTAGATGGCTTCCTATAGTGTGGCC 148

Db 50127 GTTGTGCTACTGCTCCGACAGATTTAGTTTTCATCTTAGATGGCTTCCTATAGTGTGGCC 50068

Qy 149 CAGAAACTTTGAAATAGTGAAGAGTGGCTTGTCAATATCAGAAACTTTGACATAG 208

Db 50067 CAGAAACTTTGAAATAGTGAAGAGTGGCTTGTCAATATCAGAAACTTTGACATAG 50005



```
Db 349 GGGATCCCTGGGGCTTCTGGGCTCAAGGGAGAACACAGGAGCAACGGTTCCCGAGGAGAA 408
Qy 2044 CCAGGATACATGGGTTTACCGGGATTTCAAGGAAAAAGGGGACAAAGGAAATCAAGGT 2103
Db 409 CCAGGATACATGGGTTTACCGGGATTTCAAGGAAAAAGGGGACAAAGGAAATCAAGGT 468
Qy 2104 GAAAAAGGTATTCAAGGTCAAAGGGAGAAATGGAAGACAGGGAAATTCAGGGCAACAG 2163
Db 469 GAAAAAGGTATTCAAGGTCAAAGGGAGAAATGGAAGACAGGGAAATTCAGGGCAACAG 528
Qy 2164 GGAATTCAGGCCCATCATGGTGCAAAAGCAGAGAGAGGTGAAAAGGGAGAACTTGGTGTG 2223
Db 529 GGAATTCAGGCCCATCATGGTGCAAAAG--GAGAGNGGTGAAAAGGGAGAACCTTGGTGTG 586
Qy 2224 CGAGGTGCCATTCAGATCAAAAGGAGAAATTCGGGGTGGATGGCTTGATGGGCGCCGACAGT 2283
Db 587 CGAGGTGCCATTCAGATCAAAAGGAGAAATTCGGGGTGGATGGCTTGATGGGCGCCGACAGT 646
Qy 2284 CCTAAGGGCAACCTCGGGATCCAGGTCTCTCAGGACACCCCGACAGGTTCGGATGGGAAGCC 2343
Db 647 CCTAAGGGCAACCTCGGGATCCAGGTCTCTCAGGACACCCCGACAGGTTCGGATGGGAAGCC 706
Qy 2344 GGAAGAGAGTTTTCAGAAACAATTTATTCAGCAAGTTTGCACAGATGTATTAAGAGCCAG 2403
Db 707 GGAAGAGAGTTTTCAGAAACAATTTATTCAGCAAGTTTGCACAGATGTATTAAGAGCCAG 766
Qy 2404 CTACCACTCTTACTTCAGAGTGGAGAAATTCAGAAATTCAGATTCATTCCTGTCACCAAT 2463
Db 767 CTACCACTCTTACTTCAGAGTGGAGAAATTCAGAAATTCAGATTCATTCCTGTCACCAAT 826
Qy 2464 GGCTCCCGGGTATTCCTGGGCCACCTGGTCCGATAGGCCCGACAGGGTCCAGAGGATTA 2523
Db 827 GGCTCCCGGGTATTCCTGGGCCACCTGGTCCGATAGGCCCGACAGGGTCCAGAGGATTA 886
Qy 2524 CCTGGTTTCCAGAGAGAGATGGTGTCTCGATAGTGGTGTCTCGATAGGTCCTGGAGTCCAGGT 2583
Db 887 CCTGGTTTCCAGAGAGAGATGGTGTCTCGATAGTGGTGTCTCGATAGGTCCTGGAGTCCAGGT 946
Qy 2584 GTCAGAGGATTAAGAGCCCTTACCCAGGAGAAATGGGCAAAAGGGAGCAAGGGTTTGGG 2643
Db 947 GTCAGAGGATTAAGAGCCCTTACCCAGGAGAAATGGGCAAAAGGGAGCAAGGGTTTGGG 1006
Qy 2644 TATCTCGAGAGCAAGGTCTCTCGTTCCTCGGATCCAGAGTCCAGAGGCGCTCTCGAATAGC 2703
Db 1007 TATCTCGAGAGCAAGGTCTCTCGTTCCTCGGATCCAGAGTCCAGAGGCGCTCTCGAATAGC 1066
Qy 2704 AAAGAGGTCCTCCAGAGACCCAGGTCTCTCGGATCCAGAGTCCAGAGTCCAGAGGAGCC 2763
Db 1067 AAAGAGGTCCTCCAGAGACCCAGGTCTCTCGGATCCAGAGTCCAGAGTCCAGAGGAGCC 1126
Qy 2764 GGAATCCAGGGCAACAGGCGCCCGCAGGATCCAGGATCCAGGATCCAGGATCCAGGAT 2823
Db 1127 GGAATCCAGGGCAACAGGCGCCCGCAGGATCCAGGATCCAGGATCCAGGATCCAGGAT 1186
Qy 2824 ATTGCCAGAGAGATCCGTTTCAGAAAAGGACCAAACTATTAG 2865
Db 1187 ATTGCCAGAGAGATCCGTTTCAGAAAAGGACCAAACTATTAG 1228

RESULT 8
AX552238 LOCUS 1115 bp DNA linear PAT 27-NOV-2002
DEFINITION Sequence 147 from Patent WO0162927.
ACCESSION AX552238
VERSION AX552238.1 GI:25896476
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Banville,S.C., Greenawalt,L.B., Lincoln,S.E., Stockdreher,T.K.,
```

```
Amshay,S., Chang,S.C., Chen,W., D'Sa,S.A., Dam,T.C., Liu,T.F.,
Rosen,B.H., Russo,F.D., Spiro,P.A., Bradley,D.L., Chen,A.,
Cohen,H.J., Daffo,A., Daniels,S.E., Dufour,G.E., Flores,V.,
Fong,W.T., Hodgson,D.M., Jackson,S., Jones,A.L., Panzer,S.,
Roseberry,A.M., Shah,P., Wright,R.J., Yap,P.E., Yu,J.Y.,
Bratcher,S.R., Chalup,M.S., Dahl,C.R. and Hillman,J.L.
Polypeptides and corresponding polynucleotides for diagnostics and
therapeutics
JOURNAL Patent: WO 0162927-A 147 30-AUG-2001;
Incyte Genomics, Inc. (US)
FEATURES
Location/Qualifiers
Source 1..115
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/note="Incyte ID No: LG:120744.1:2000MAY19"
BASE COUNT 344 a 215 c 259 g 296 t 1 others
ORIGIN
Query Match 26.1%; Score 749; DB 6; Length 1115;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 919; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
Qy 1 ATGGCTCACTATATACATTTCTCTGATGGTTTGGTGTCTCTCAGAAATCTGTG 60
Db 196 ATGGCTCACTATATACATTTCTCTGATGGTTTGGTGTCTCTCAGAAATCTGTG 255
Qy 61 TTAGCTGAAGATGGGGAAGTAAGATCAAGTTGCTGCTACTGCTCCGACAGATTTAGTTTC 120
Db 256 TTAGCTGAAGATGGGGAAGTAAGATCAAGTTGCTGCTACTGCTCCGACAGATTTAGTTTC 315
Qy 126 ATCTTAGATGGCTCTTATAGTGTGGCCACAGAAACTTTGAAATAGTGAATAAGTGCCT 180
Db 316 ATCTTAGATGGCTCTTATAGTGTGGCCACAGAAACTTTGAAATAGTGAATAAGTGCCT 375
Qy 181 GTCATATACAAATAACTTTGACATAGGCGCGAGTTTATTCAGTGGAGTGTCAA 240
Db 376 GTCATATACAAATAACTTTGACATAGGCGCGAGTTTATTCAGTGGAGTGTCAA 435
Qy 241 TATAGTGACTACCTCTGCTGAGATTCCTCTCGAAGCTATGATTCAGGAGAACATTG 300
Db 436 TATAGTGACTACCTCTGCTGAGATTCCTCTCGAAGCTATGATTCAGGAGAACATTG 495
Qy 301 AGCGAGCAGTGGAAATCCATCTCTACTTAGAGGAAACACAAAGACAGGAGAGGCCATC 360
Db 496 AGCGAGCAGTGGAAATCCATCTCTACTTAGAGGAAACACAAAGACAGGAGAGGCCATC 555
Qy 361 CAGTTTGGCTCGATTACCTTTTGGCAAGTCTTCTGACTTCTGACTAAGATAGCAGTG 420
Db 556 CAGTTTGGCTCGATTACCTTTTGGCAAGTCTTCTGACTAAGATAGCAGTG 675
Qy 421 GTACTTAGGATGGCAATCCCAAGATGACGTCAGGATCGAGTCAAGCAGCAAGAT 480
Db 616 GTACTTAGGATGGCAATCCCAAGATGACGTCAGGATCGAGTCAAGCAGCAAGAT 675
Qy 481 AGTAAGATAACATTTATTTGCTATTGGTGTTCGTTTCAGAAACAGAGATGCCGAATAGA 540
Db 676 AGTAAGATAACATTTATTTGCTATTGGTGTTCGTTTCAGAAACAGAGATGCCGAATAGA 735
Qy 541 GCTATTGCCAACAGCCCTTCTGCTACTTATGTTTATGTTGGAAGCTATATTGCAATA 600
Db 736 GCTATTGCCAACAGCCCTTCTGCTACTTATGTTTATGTTGGAAGCTATATTGCAATA 795
Qy 601 TCCAAAATAAGGGAAGTGTGTAAGCAGAAACTTTGTGAAGAATCTGCTCTCAACACGA 660
Db 796 TCCAAAATAAGGGAAGTGTGTAAGCAGAAACTTTGTGAAGAATCTGCTCTCAACACGA 855
Qy 662 ATTCCAGTGGCAGCTCGTGATGAAGGGGATTTGATATTTCTTTAGGTTTAGATGTAAT 720
Db 856 ATTCCAGTGGCAGCTCGTGATGAAGGGGATTTGATATTTCTTTAGGTTTAGATGTAAT 914
Qy 721 AAAAAGGTTAGAAAAGATACAGCTTTCACCAAAAAGATATAAGAGATATGAAGTAACA 780
```

source	1. .1169	Db	1096	TTAACTATTGATGGAGGCGCACAAATAGCAGTTCCTTAATGCTGGGACAAATCTTA	1155
	/organism="Homo sapiens"				
	/mol_type="genomic DNA"				
	/db_xref="taxon:9606"				
	/note="Incyte ID No: LG:120744.1:2000SEP08"				
BASE COUNT	364 a 224 c 267 g 314 t				
ORIGIN					
Query Match	32.2%; Score 923; DB 6; Length 1169;				
Best Local Similarity	99.9%; Pred. No. 0;				
Matches	973; Conservative 0; Mismatches 1; Indels 0; Gaps 0;				
Qy	1 ATGGCTCAGTATATTAATTTCTCTGATGTTTGGTCTGCTCTTTCAGAAATCTGTG	60			
Db	196 ATGGCTCAGTATATTAATTTCTCTGATGTTTGGTCTGCTCTTTCAGAAATCTGTG	255			
Qy	61 TTAGCTGAGATGGGAGTAAATCAAGTTCGTTGCTGCTCCGACAGATTTAGTTTC	120			
Db	256 TTAGCTGAGATGGGAGTAAATCAAGTTCGTTGCTGCTCCGACAGATTTAGTTTC	315			
Qy	121 ATCTAGATGGCTCTTAGTGTGGCCAGAAAATTTGAATAGTGAAGAAGTGGTT	180			
Db	316 ATCTAGATGGCTCTTAGTGTGGCCAGAAAATTTGAATAGTGAAGAAGTGGTT	375			
Qy	181 GTCAATATCAAAAATCTTTGACATAGGCCCAAGTTTATTCAGTTGAGTGTCTCA	240			
Db	376 GTCAATATCAAAAATCTTTGACATAGGCCCAAGTTTATTCAGTTGAGTGTCTCA	435			
Qy	241 TATAGTGAATACCTGTGCTGGAGATTCCTCTCGAAGCTATGATTCAGAGAACTTG	300			
Db	436 TATAGTGAATACCTGTGCTGGAGATTCCTCTCGAAGCTATGATTCAGAGAACTTG	495			
Qy	301 ACGCAGCAGTGGATCCATCTCTAGTAGGAGGAAACACAAAGACAGGAGCCATC	360			
Db	496 ACGCAGCAGTGGATCCATCTCTAGTAGGAGGAAACACAAAGACAGGAGCCATC	555			
Qy	361 CAGTTTGGCTCGATTAACCTTTTGGCAAGTCTCTCAGATTTCTGCTAAGTAGCAGTG	420			
Db	556 CAGTTTGGCTCGATTAACCTTTTGGCAAGTCTCTCAGATTTCTGCTAAGTAGCAGTG	615			
Qy	421 GTACTACGGATGCAATCCAGATGAGTCAAGATGACGTCACGTCACGACACAGAGAT	480			
Db	616 GTACTACGGATGCAATCCAGATGAGTCAAGATGACGTCACGTCACGACACAGAGAT	675			
Qy	481 AGTAAGATAACATTATTTGCTATTTGGTGTGGTTGAGTTCAGAAACAGAGATCCGA	540			
Db	676 AGTAAGATAACATTATTTGCTATTTGGTGTGGTTGAGTTCAGAAACAGAGATCCGA	735			
Qy	541 GCTATTGCCAAGCCCTTCGCTACTTATGTTGTTTATGTTGGAAGACTATATTCGATA	600			
Db	736 GCTATTGCCAAGCCCTTCGCTACTTATGTTGTTTATGTTGGAAGACTATATTCGATA	795			
Qy	601 TCCAAAATAAGGGAAGTGAAGACAGAAACTTTGTGAAGAATCTCTGTCTCAACACGA	660			
Db	796 TCCAAAATAAGGGAAGTGAAGACAGAAACTTTGTGAAGAATCTCTGTCTCAACACGA	855			
Qy	661 ATCCAGTGGAGCTCGTGAAGAGGATTTGATATCTTTTAGTTAGATTAAT	720			
Db	856 ATCCAGTGGAGCTCGTGAAGAGGATTTGATATCTTTTAGTTAGATTAAT	915			
Qy	721 AAAAAGGTTAAGAAAGAAATACAGCTTTTACCAAAAAGATAAAGATATGAAGTAA	780			
Db	916 AAAAAGGTTAAGAAAGAAATACAGCTTTTACCAAAAAGATAAAGATATGAAGTAA	975			
Qy	781 TCAAAAGTTGATTTATCAGAACTCAGAGCAATGTTTCCAGAGGCTCTCTCCATCA	840			
Db	976 TCAAAAGTTGATTTATCAGAACTCAGAGCAATGTTTCCAGAGGCTCTCTCCATCA	1035			
Qy	841 TATGTTATTTGCTCTACTCAAAAGTTTAAAGTCAAGAAATTTGGGATTTATGGAA	900			
Db	1036 TATGTTATTTGCTCTACTCAAAAGTTTAAAGTCAAGAAATTTGGGATTTATGGAA	1095			
Qy	901 TTAACATATTGGAAGGCCCAAAATAGCAGTTACCTTAATGCTGGACAAATCTTA	960			
LOCUS	AK096444				
DEFINITION	Homo sapiens cDNA FL35125 f1s, clone NTONG2607034.				
ACCESSION	AK096444				
VERSION	AK096444.1 GI:21755943				
KEYWORDS	oligo capping; f1s (full insert sequence).				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	1				
AUTHORS	Okamoto, M., Tanase, T., Imase, N., Takeuchi, K., Arita, M., Musashino, K., Yuki, H., Hara, H., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Sano, Y., Kawai-Hiro, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Wagatsuma, M., Murakawa, K., Kanehori, K., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahara, K., Masuhara, Y., Nagai, K. and .Sogai, T.				
TITLE	NEDO human cDNA sequencing project				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 2295)				
AUTHORS	Isogai, T. and Yamamoto, J.				
TITLE	Direct Submission				
JOURNAL	Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan				
COMMENT	IE-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986; NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology (RAB); cDNA library construction; Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing; RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing; HRI and RAB; annotation: HRI and RAB.				
FEATURES	Location/Qualifiers				
source	1. .2295				
	/organism="Homo sapiens"				
	/mol_type="mRNA"				
	/db_xref="taxon:9606"				
	/clone="NTONG2007034"				
	/issue_type="longue"				
	/clone_lib="NTONG2"				
	/note="cloning vector: pME:8SFL3"				
BASE COUNT	683 a 427 c 565 g 620 t				
ORIGIN					
Query Match	30.7%; Score 880; DB 9; Length 2295;				
Best Local Similarity	99.8%; Pred. No. 0;				
Matches	1060; Conservative 0; Mismatches 0; Indels 2; Gaps 2;				
Qy	1804 GGATCCAGGATTTCTTGGAAACCGAGGATTAATGGCCAAAAGGAGGAAATGGG	1863			
Db	169 GGATCCAGGATTTCTTGGAAACCGAGGATTAATGGCCAAAAGGAGGAAATGGG	228			
Qy	1864 CAGGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	1923			
Db	229 CAGGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	288			
Qy	1924 TCACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	1983			
Db	289 TCACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	348			
Qy	1984 GGATCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	2043			

```
Db 2603 ATAGAGCCAGCTACCACTCTTACTCAGAGTGAAGAAATTAGAAATTGTGATCATTC 2662
QY 2452 CTGTCCCAACATGGCTCCCGGGTATTCTCGGCCACCTGGTCCGATAGGCCACAGAGGT 2511
Db 2663 CTGTCCCAACATGGCTCCCGGGTATTCTCGGCCACCTGGTCCGATAGGCCACAGAGGT 2722
QY 2512 CCAGAGGATACCTGGTTTCCAGAGAGAGATGGTGTCTCTGATTTAGTGGGTGTCCT 2571
Db 2723 CCAGAGGATACCTGGTTTCCAGAGAGAGATGGTGTCTCTGATTTAGTGGGTGTCCT 2782
QY 2572 GCAGCTCCAGGTTGTCAGAGGATTAAAGGCTTACCAGGAAGAAATGGGAAAAAGGAGC 2631
Db 2783 GCAGCTCCAGGTTGTCAGAGGATTAAAGGCTTACCAGGAAGAAATGGGAAAAAGGAGC 2842
QY 2632 CAAGGTTTGGTATCTCGAGAGACAGGTCTCTCTGTCCTCCAGGTCCAGAGGGCCCT 2691
Db 2843 CAAGGTTTGGTATCTCGAGAGACAGGTCTCTCTGTCCTCCAGGTCCAGAGGGCCCT 2902
QY 2692 CCTGGAATAGCAAGAGAGGTCTCTCAGAGACCCAGGTCTCTCTGTCCTCCAGATGGAGAC 2751
Db 2903 CCTGGAATAGCAAGAGAGGTCTCTCAGAGACCCAGGTCTCTCTGTCCTCCAGATGGAGAC 2962
QY 2752 CATGAAAACCTGGATCCAGGGCAACGAGCCCCCAGGCATCTCGAGCCCATCTA 2811
Db 2963 CATGAAAACCTGGATCCAGGGCAACGAGCCCCCAGGCATCTCGAGCCCATCTA 3022
QY 2812 TGTATTAGTGAATTCGAGAGAGATCCGTTTCAGAAAAGGACCAAACTATTAG 2965
Db 3023 TGTATTAGTGAATTCGAGAGAGATCCGTTTCAGAAAAGGACCAAACTATTAG 3076

RESULT 5
AX430410
LOCUS 1184 bp DNA linear PAT 28-JUN-2002
DEFINITION Sequence 192 from Patent WO0240715.
ACCESSION AX430410
VERSION AX430410.1 GI:21655774
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1
AUTHORS Chalup,M.S., Altus,C.M., Lincoln,S.E., Dufour,G.E. and Jackson,S.
TITLE Molecules for disease detection and treatment
JOURNAL Patent: WO 0240715-A 192 23-MAY-2002;
INCYTE GENOMICS INC (US)
FEATURES
Source
1. 1184
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/note="Incyte ID No: LI:108819.1:2000SEP08"
BASE COUNT 368 a 227 c 272 g 317 t
ORIGIN

Query Match 34.0%; Score 974; DB 6; Length 1184;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 974; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCTCACTATATACATTTCTCTGATGTTTGGTCTCTCTTCAGAAATCTGTG 60
Db 211 ATGGCTCACTATATACATTTCTCTGATGTTTGGTCTCTCTTCAGAAATCTGTG 270
QY 61 TTAGCTGAAGATGGGAAGTAAAGTCAAGTCTCTGATGTTTGGTCTCTTCAGAAATCTGTG 120
Db 271 TTAGCTGAAGATGGGAAGTAAAGTCAAGTCTCTGATGTTTGGTCTCTTCAGAAATCTGTG 330
QY 121 ATCTTAGATGGCTCTTATAGTGTGGCCAGAAAATCTTGAATAGTGAAGAGTGGCTT 180
Db 331 ATCTTAGATGGCTCTTATAGTGTGGCCAGAAAATCTTGAATAGTGAAGAGTGGCTT 390
QY 181 GTCAATATCAAAAAAATTTTGACATAGGGCCGAAATTTATTCAAGTTGGAGTGGTTCAA 240
```

```
Db 391 GTCAATATCAAAAAAATTTTGACATAGGGCCGAAAGTTATTCAAGTTGGAGTGGTTCAA 450
QY 241 TATAGTGAATACCTCTGTGCTGGAGATTCCTCTCGAAGCTATGATTTCAGAGAAATTTG 300
Db 451 TATAGTGAATACCTCTGTGCTGGAGATTCCTCTCGAAGCTATGATTTCAGAGAAATTTG 510
QY 301 ACGGAGAGTGGAAATCCATACCTACTTACTTAGGAGGAAACACAAAGACAGGGAAGGCATC 360
Db 511 ACGGAGAGTGGAAATCCATACCTACTTACTTAGGAGGAAACACAAAGACAGGGAAGGCATC 570
QY 361 CAGTTTGGCTCGATTAACCTTTTGGCAAGTCTCTCAGATTTCTGACTTAAGATAGCAGTG 420
Db 571 CAGTTTGGCTCGATTAACCTTTTGGCAAGTCTCTCAGATTTCTGACTTAAGATAGCAGTG 630
QY 421 GTACTTACGAGTGGCAAAATCCCAAGATGACGTCAAGGATGACGTCAAGCAGCAGAGAT 480
Db 631 GTACTTACGAGTGGCAAAATCCCAAGATGACGTCAAGGATGACGTCAAGCAGCAGAGAT 690
QY 481 AGTAAATTAACATTAATTTGCTATTGGTGGTTCAGAAACAGAGATGCCAACTTACA 540
Db 691 AGTAAATTAACATTAATTTGCTATTGGTGGTTCAGAAACAGAGATGCCAACTTACA 750
QY 541 GCTATTGCCAAAGCCCTCTGCTACTTATGTGTTTATGTGGAAGACTATATTCGAAATA 600
Db 751 GCTATTGCCAAAGCCCTCTGCTACTTATGTGTTTATGTGGAAGACTATATTCGAAATA 810
QY 601 TCCAAATTAAGGAGTGAATGACGAGCAAAACTTTGCAAGAAATCTCTGTCGCAACAGA 660
Db 811 TCCAAATTAAGGAGTGAATGACGAGCAAAACTTTGCAAGAAATCTCTGTCGCAACAGA 870
QY 661 ATCCAGTGGCAGCTCTGATGAAGGGGATTTGATATCTCTTTAGGTTTAGATGTAAT 720
Db 871 ATCCAGTGGCAGCTCTGATGAAGGGGATTTGATATCTCTTTAGGTTTAGATGTAAT 930
QY 721 AAAAGGTTAAGAAAGATACAGCTTTCACCAAAAAGATAAAGAGATAGAGTAACA 780
Db 931 AAAAGGTTAAGAAAGATACAGCTTTCACCAAAAAGATAAAGAGATAGAGTAACA 990
QY 781 TCAAAAGTTGATTTATCAGAACTCACAAGCAATGTTTCCCAAGAGGTCTCTCTCCATCA 840
Db 991 TCAAAAGTTGATTTATCAGAACTCACAAGCAATGTTTCCCAAGAGGTCTCTCTCCATCA 1050
QY 841 TATGATTTGTGCTACTCTCAAGATTTAAAGTCAAGAAATTTGGGATTTATGGAGATA 900
Db 1051 TATGATTTGTGCTACTCTCAAGATTTAAAGTCAAGAAATTTGGGATTTATGGAGATA 1110
QY 901 TTAATATTGATGAAGGCCACAAATAGCAGTTTACCTTAAATGGTGGACAAAATCTTA 960
Db 1111 TTAATATTGATGAAGGCCACAAATAGCAGTTTACCTTAAATGGTGGACAAAATCTTA 1170
QY 961 TTAATTTACAACAAC 974
Db 1171 TTAATTTACAACAAC 1184

RESULT 6
AX430281
LOCUS 1169 bp DNA linear PAT 28-JUN-2002
DEFINITION Sequence 63 from Patent WO0240715.
ACCESSION AX430281
VERSION AX430281.1 GI:21655645
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1
AUTHORS Chalup,M.S., Altus,C.M., Lincoln,S.E., Dufour,G.E. and Jackson,S.
TITLE Molecules for disease detection and treatment
JOURNAL Patent: WO 0240715-A 63 23-MAY-2002;
INCYTE GENOMICS INC (US)
FEATURES
Location/Qualifiers
```

QY	241	TATAGTGACTACCCCTGTGCTGGAGATTCTCTCTCGAAAGCTATGATTGAGAGAAACATTG	300
DB	443	TATAGTGACTACCCCTGTGCTGGAGATTCTCTCTCGAAAGCTATGATTGAGAGAAACATTG	502
QY	301	ACGCACAGTGGAAATCCATACTCTACTTAGGAGGAACACAAAGACAGGGAAGGCCATC	360
DB	503	ACGCACAGTGGAAATCCATACTCTACTTAGGAGGAACACAAAGACAGGGAAGGCCATC	562
QY	361	CAGTTTGGCTCGATTACCTTTTTTCCAAAGTCTCAGATTCTTGACTAAGATAGCAGTG	420
DB	563	CAGTTTGGCTCGATTACCTTTTTTGCACAAAGTCTCAGATTCTTGACTAAGATAGCAGTG	622
QY	421	GTACTTACGGATGCGAAATCCACAGATGACGTCAAGATGCAGCTCAAGCAGACAGAGAT	480
DB	623	GTACTTACGGATGCGAAATCCACAGATGACGTCAAGATGCAGCTCAAGCAGACAGAGAT	682
QY	481	AGTAAGATAACATTATTTGCTATTGCTGTGGTTGCGTTTCAGAAAACAGAAAGATGCGAACTTGA	540
DB	683	AGTAAGATAACATTATTTGCTATTGCTGTGGTTTCAGAAAACAGAAAGATGCGAACTTGA	742
QY	541	GCTATTGCCAACAGCCCTTCGTCTACTTATGTGTGTTTTATGTGGAAGACTATATTGCAATA	600
DB	743	GCTATTGCCAACAGCCCTTCGTCTACTTATGTGTGTTTTATGTGGAAGACTATATTGCAATA	802
QY	601	TCCAAAATAAGGGGAAGTGCATCAAGCAGAAAATTTGTGAAGAAATCTGTCTGTCCAAACAGA	660
DB	803	TCCAAAATAAGGGGAAGTGCATCAAGCAGAAAATTTGTGAAGAAATCTGTCTGTCCAAACAGA	862
QY	661	ATTCAGTGGGAGCTCGTGATGAAAGGGGATTTGATATTTCTTTAGGTTAGATGTAAAT	720
DB	863	ATTCAGTGGGAGCTCGTGATGAAAGGGGATTTGATATTTCTTTAGGTTAGATGTAAAT	922
QY	721	AAAAAGGTTAAGAAAAGAAATACAGCTTTTACCAAAAAGATAAAAGCATATGAAGTAAACA	780
DB	923	AAAAAGGTTAAGAAAAGAAATACAGCTTTTACCAAAAAGATAAAAGCATATGAAGTAAACA	982
QY	781	TCAAAAGTTGATTTATCAGAACTCACAAGGAATGTTTTCCAGAAAGGTCTTCTCCATCA	840
DB	983	TCAAAAGTTGATTTATCAGAACTCACAAGGAATGTTTTCCAGAAAGGTCTTCTCCATCA	1042
QY	841	TATGTTATTTGTCTACTCCTCAAGATTTTAAAGTCAAGAAAATTTGCGAATTTATGGAGAAAT	900
DB	1043	TATGTTATTTGTCTACTCCTCAAGATTTTAAAGTCAAGAAAATTTGCGAATTTATGGAGAAAT	1102
QY	901	TTAACTATTGATGAAGGCCACAAATAGCAGTTACCTTTAAATGCTGTGGAACAAATCTTA	960
DB	1103	TTAACTATTGATGAAGGCCACAAATAGCAGTTACCTTTAAATGCTGTGGAACAAATCTTA	1162
QY	961	TTATTTACAACAAACAGCGTAATTAATGGCTCACAAGTGTACCTTTGCTTACCTCAA	1020
DB	1163	TTATTTACAACAAACAGCGTAATTAATGGCTCACAAGTGTACCTTTGCTTACCTCAA	1222
QY	1021	GTTAAGACGTTGTTTGATGAAGGTGGCACCAAAATTCGTCTCTAGTAAACAGAACAAAGAT	1080
DB	1223	GTTAAGACGTTGTTTGATGAAGGTGGCACCAAAATTCGTCTCTAGTAAACAGAACAAAGAT	1282
QY	1081	GTGACTTTGATATATGTAGCCACCAAAATTCGAAACAGCCCTTACATCCAGTTTAGGG	1140
DB	1283	GTGACTTTGATATATGTAGCCACCAAAATTCGAAACAGCCCTTACATCCAGTTTAGGG	1342
QY	1141	ATCTTGATCAATGGGCAAAACCAAAATTTGAAATTAATCTGAAAAGGAAGAACTGTTCAG	1200
DB	1343	ATCTTGATCAATGGGCAAAACCAAAATTTGAAATTAATCTGAAAAGGAAGAACTGTTCAG	1402
QY	1201	TTTTCATGTCGAAAAGTTGCGAATCTACTGTGACCCAGAAACAGAACACCCGGAGACAGCA	1260
DB	1403	TTTTCATGTCGAAAAGTTGCGAATCTACTGTGACCCAGAAACAGAACACCCGGAGACAGCA	1462
QY	1261	TGTGAGATTCTCGGATTT-----TGCCTTAATGTCCTCCAGTGATGTAGGTTCAACT	1311
DB	1463	TGTGAGATTCTCGGATTTAAATGGAGAGTGCCCTTAATGGTCCAGTGATGTAGGTTCAACT	1522
QY	1312	CCAGCTCCCTGATTGTTGCTCCCGGAAAACACAGACTTCAAGGCCCCCAAGGTGACCTT	1371

```
QY 2092 GGAATCAAGGTGAAAGAGTATTGAGGTCAAAAGGAGAGAAATGGAACACAGCGAATT 2151
DB 2303 GGAATCAAGGTGAAAGAGTATTGAGGTCAAAAGGAGAGAAATGGAACACAGCGAATT 2362
QY 2152 CCAGGCAACAGGGAATTCAGAGCCATCATGTCGAAAGGAGAGAGAGTGAAGGGA 2211
DB 2363 CCAGGCAACAGGGAATTCAGAGCCATCATGTCGAAAGGAGAGAGAGTGAAGGGA 2422
QY 2212 GAACCTGTGTCAGAGTGCCATTTGGATCAAAAGGAGAACTCGGGTGGATGCTTGATG 2271
DB 2422 GAACCTGTGTCAGAGTGCCATTTGGATCAAAAGGAGAACTCGGGTGGATGCTTGATG 2482
QY 2272 GGGCCCGCAGGTCTTAAGGGGCAACCTGGGGATCCAGTCTCCAGGACCCCGCAGGTTG 2331
DB 2482 GGGCCCGCAGGTCTTAAGGGGCAACCTGGGGATCCAGTCTCCAGGACCCCGCAGGTTG 2542
QY 2332 GATGGAGCCCGGAGAGAGTTTTCAGACAAATTTATTCAGACAGTTTTCACAGATGTA 2391
DB 2543 GATGGAGCCCGGAGAGAGTTTTCAGACAAATTTATTCAGACAGTTTTCACAGATGTA 2602
QY 2392 ATAAGAGCCCGCAGTACCAAGTCTTACTTCAGAGTGGAGAAATAGAAATTTGATCATTCG 2451
DB 2602 ATAAGAGCCCGCAGTACCAAGTCTTACTTCAGAGTGGAGAAATAGAAATTTGATCATTCG 2662
QY 2452 CTGTCCCAACATGCTCCCGGGTATTCTCGGGCCACCTGGTCCGATAGCCCGCAGAGGT 2511
DB 2662 CTGTCCCAACATGCTCCCGGGTATTCTCGGGCCACCTGGTCCGATAGCCCGCAGAGGT 2722
QY 2512 CCCAGAGATTACCTGGTTTCCAGAGGAGATGTTCTCTCGATAGTGGTGGTCCCT 2571
DB 2722 CCCAGAGATTACCTGGTTTCCAGAGGAGATGTTCTCTCGATAGTGGTGGTCCCT 2782
QY 2572 GGAACCTCCAGGTGTCAGAGGATTAAGAGCCCTACAGGAGAAATGGGGGAAAGGAGC 2631
DB 2782 GGAACCTCCAGGTGTCAGAGGATTAAGAGCCCTACAGGAGAAATGGGGGAAAGGAGC 2842
QY 2632 CAAAGGTTGGGTATCTCGAGACACAGTCTCTCTGTCCTCCAGGTCAGAGGCGCT 2691
DB 2842 CAAAGGTTGGGTATCTCGAGACACAGTCTCTCTGTCCTCCAGGTCAGAGGCGCT 2902
QY 2692 CTGGATTAAGCAAGAGAGTCTCTCCAGGACCCAGTCTCTCTGTCCTCCAGGTCAGAGG 2751
DB 2902 CTGGATTAAGCAAGAGAGTCTCTCCAGGACCCAGTCTCTCTGTCCTCCAGGTCAGAGG 2962
QY 2752 CATGAAACCTGGAATCAAGGCAACAGGCCCCCAGGATCTCGGACCCATCACTA 2811
DB 2962 CATGAAACCTGGAATCAAGGCAACAGGCCCCCAGGATCTCGGACCCATCACTA 3022
QY 2812 TGTTTGTGTAATTCAGAGAGATCCGTTTCAGAAAGGACCAAACTATTAG 2865
DB 3023 TGTTTGTGTAATTCAGAGAGATCCGTTTCAGAAAGGACCAAACTATTAG 3076

RESULT 4
HSM801597 4160 bp mRNA linear PRI 20-MAR-2002
LOCUS HSM801597
DEFINITION Homo sapiens mRNA; cDNA DKFZp564B052 (from clone DKFZp564B052);
complete cds.
ACCESSION AL136624
VERSION AL136624.1 GI:12052773
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Duesterhoeft.A., Lauber.J., Mewes,H.W., Gassenhuber.J. and
Wiemann,S.
Direct Submission
Submitted (12-MAR-2002) MIPS, Am Klopfer:spitz 18a, D-82152
Martinsried, GERMANY
COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wie-ann@dkfz-heidelberg.de;
```

```
sequenced by Qiagen (Hilden/Germany) within the cDNA sequencing
consortium of the German Genome Project.
This clone (DKFZp564B052) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clonedrpd.de Further
information about the clone and the sequencing project is available
at http://mips.gsf.de/proj/cDNA/.

FEATURES             Location/Qualifiers
     source            1..4160
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /map="6"
                     /clone="DKFZp564B052"
                     /tissue_type="brain"
                     /clone_lib="564 (synonym: hfbz2). Vector pAMP1; host
                     X1-2blue; sites NotI + SalI;"
                     /dev_stage="fetal"
     gene              1..4160
                     /gene="DKFZp564B052"
     CDS                203..3076
                     /gene="DKFZp564B052"
                     /note="DJ708F5.1, differences to genmodel, similarity to
                     collagen proteins"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAB66559.1"
                     /db_xref="GI:12052774"
                     /db_xref="SPTREMBL:Q9H0V3"
                     /translation="MAHITFLCVLVLQNSVLAEDGEVRSRSCPTDLDVILDG
                     SYSVGFNSIVKLVNITNFEDIGPKFIQGVWQYSDYVLEIPGYSYDSGEHJTA
                     AVES:LYLNGNTKTGAIPALDYLFDKSRFLTKIAVLTDGKSQDDVKDAQAARD
                     SKITFLAIGVSTEDAEALAIANKPSSTVYFVEDYIAISK:REVNMKQLCBESVCP
                     TRIPVAARDERFDILLDVNVKVKRIOLSPKIKGYEVTSKVLSELTNSVPEEG
                     /PSYFVSTQREKVKXIMDLRLIT:DGSPQ:AVTLNGVDK:LLFTTSVINGSOVY
                     TPANPVKTLFDEGHOIRLLVTEQSVTLX:IDSOQ:ENKLPHPGLGILLINGTOIGKY
                     SKGETVQFVDKLRUYCDPEQNNRETACIPGNECLNGSPDSVSTFAP:CLCPGKX
                     PGLQPKGDVLPNGPYPQPGQDQKPGYAGT:PGVPGSPGI:QGARGLPGYKXGP
                     GRGCKDGLDLPGLHMPGSGMGAKGDKSPGPGYKKGAKGKNGAGFPGLPG
                     PAGEPGRHGDG:WGSYPGFKGAGSPGAPQDQGTGREGIPGPGKRNGLMGKGLGELP
                     PGQCKGKAPGMPLMGNSGSPQPGT:PGSKSGKGRPGI:QMPGAGSLGKEPGATGSP
                     GERVMGLPGI:QKKGKNGKNGEKIGCGKNGKNGRGI:PGOOGI:QSHGKNGKGERGK
                     EPGVRGALGSKGSGVDGLMGAPGKQPGQDQPGPQPPGLDQKPREFSEQIRQVCT
                     DVITRAQ:LVLLQSGRI:RNCHLUSQMGSPGIPGPPGPI:GEPGPG:JPLPGRDGVPL
                     VYVPRFVGRGLKGLPGRNGEKSGQSGFVGEQGPFGPGPGPGI:SKGEGPGPGGL
                     PGKGDHGXFG:QQQPGPGI:CDPSLCFSV:ASRDPFRKGPVY"
     polyA_signal       4120..4125
                     /gene="DKFZp564B052"
     polyA_site         4141
                     /gene="DKFZp564B052"
BASE COUNT 1271 a 795 c 1026 g 1068 t
ORIGIN

Query Match 71.9%; Score 2061; DB 9; Length 4160;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 2861; Conservative 0; Mismatches 4; Indels 9; Gaps 1;

QY 1 ATGGCTCACTATATACATTCTCTCGATCGGTTTGTGCTGCTCTTCTTGAAATTTCTGTG 60
DB 203 ATGGCTCACTATATACATTCTCTCGATCGGTTTGTGCTGCTCTTCTTGAAATTTCTGTG 262
QY 61 TTAGCTGAAGATGGGAAGTAAGATCAAGTTGTCGTACTGCTCCGACAGATTAGTTTTC 120
DB 263 TTAGCTGAAGATGGGAAGTAAGATCAAGTTGTCGTACTGCTCCGACAGATTAGTTTTC 322
QY 121 ATCTTAGTGGCTCTTATAGTGTGTGGCCAGAAAACCTTTGAAATAGTGAAGAGTGGCTT 190
DB 323 ATCTTAGTGGCTCTTATAGTGTGTGGCCAGAAAACCTTTGAAATAGTGAAGAGTGGCTT 382
QY 131 GTCATATACAAAAAATCTTGACATAGGCCGGAAGTTTATTTCAAGTTGAGTGGTTCAA 240
DB 383 GTCATATACAAAAAATCTTGACATAGGCCGGAAGTTTATTTCAAGTTGAGTGGTTCAA 442
```

## ORIGIN

Query Match 71.9%; Score 2061; DB 6; Length 4160;  
Best Local Similarity 99.5%; Pred. No. 0;  
Matches 2861; Conservative 0; Mismatches 4; Indels 9

Cy	1	ATGGCTC	ATATATTA	CATTCTCTG	CATGGTTTTGGT	GTGCTGCTCTT	CAGAAATTTCTG	60
Db	203	ATGGCTC	ACATATATTA	CATTCTCTG	CATGGTTTTGGT	GTGCTGCTCTT	CAGAAATTTCTG	262
Qy	61	TTAGCTGA	AGATGGG	AAAGTAA	GATCAAGTTG	TCGTACTG	CTCCGACAGATTTAGTTTTC	120
Db	263	TTAGCTGA	AGATGGG	AAAGTAA	GATCAAGTTG	TCGTACTG	CTCCGACAGATTTAGTTTTC	322
Qy	121	ATCTTAGA	TGCGCTCTTA	TATAGTTT	GGCCCGA	AAACCTTT	TGAAATAGTGA	180
Db	323	ATCTTAGA	TGCGCTCTTA	TATAGTTT	GGCCCGA	AAACCTTT	TGAAATAGTGA	382
Qy	181	GTCAATAT	CAAAAAA	CTTTGACA	TAGGGCCGA	AGTTTAT	TCAAGTTGCGAGTGGTTCAA	240
Db	383	GTCAATAT	CAAAAAA	CTTTGACA	TAGGGCCGA	AGTTTAT	TCAAGTTGCGAGTGGTTCAA	442
Qy	241	TATAGTGA	CTACCTGTG	CTGGAGAT	TCCTCTCG	GAAGCTAT	GAATTCAGAGAACATTTG	300
Db	443	TATAGTGA	CTACCTGTG	CTGGAGAT	TCCTCTCG	GAAGCTAT	GAATTCAGAGAACATTTG	502
Qy	301	ACGGCAG	CAGTGG	AAATCAT	ACTCTA	CTTAGAGG	AAACAACAAGA	360
Db	503	ACGGCAG	CAGTGG	AAATCAT	ACTCTA	CTTAGAGG	AAACAACAAGA	562
Qy	361	CAGTTTGG	CTCGAAT	TCCTTTT	GCCAAGT	CTCTCAG	ATTTCTGACTA	420
Db	563	CAGTTTGG	CTCGAAT	TCCTTTT	GCCAAGT	CTCTCAG	ATTTCTGACTA	622
Qy	421	GTA	CTTACG	ATG	GC	CAATCC	CAAGATG	480
Db	623	GTA	CTTACG	ATG	GC	CAATCC	CAAGATG	682
Qy	481	AGTAAG	ATACAT	TATTTG	CTATTGG	TGTTG	TCAGAAACAG	540
Db	683	AGTAAG	ATACAT	TATTTG	CTATTGG	TGTTG	TCAGAAACAG	742
Qy	541	GCTATTG	CCAAACA	AGCCCTT	CGTCTACT	TATG	TGTTTTATG	600
Db	743	GCTATTG	CCAAACA	AGCCCTT	CGTCTACT	TATG	TGTTTTATG	802
Qy	601	TCCAAAT	TAA	GGGAAG	TGATG	AAAGCAG	NAACCTTTG	660
Db	803	TCCAAAT	TAA	GGGAAG	TGATG	AAAGCAG	NAACCTTTG	862
Qy	661	ATTCAG	TGGCAG	CTCGT	GATG	AAAGGGG	ATTTGATATCTTTT	720
Db	863	ATTCAG	TGGCAG	CTCGT	GATG	AAAGGGG	ATTTGATATCTTTT	922
Qy	721	AAA	AAAGTT	TAA	GA	AAAGAA	TAACAGCTTTT	780
Db	923	AAA	AAAGTT	TAA	GA	AAAGAA	TAACAGCTTTT	982
Qy	781	TC	AAAGTT	TGATTT	ATCAG	AACTC	ACAAGCAATGTTT	840
Db	983	TC	AAAGTT	TGATTT	ATCAG	AACTC	ACAAGCAATGTTT	1042
Qy	841	TAT	GTATTT	TG	GTCTACT	CA	AGATTTTAA	900
Db	1043	TAT	GTATTT	TG	GTCTACT	CA	AGATTTTAA	1102
Qy	901	TTA	ACTATT	TG	TGAAG	CCCA	AAATAGCAGTT	960
Db	1103	TTA	ACTATT	TG	TGAAG	CCCA	AAATAGCAGTT	1162
Qy	961	TTATTTA	CA	AAAC	CAG	CGTAA	TTAATGGCT	1020
Db	1163	TTATTTA	CA	AAAC	CAG	CGTAA	TTAATGGCT	1222

Qy	1021	GTTAAGAGCTGTGTTGATGAAGCTGGCAACCAAAATTCGTCCTCTTAGTAAACAGAACAGAT	1080
Dp	1223	GTTAAGACGCTGTGTTGATGAAGCTGGCAACCAAAATTCGTCCTCTTAGTAAACAGAACAGAT	1282
Qy	1081	GTGACTTTGCTATATATGATGACCAACAATTTGNAAGACAGCCCTACATCCAGTTTATGGG	1140
Dp	1283	GTGACTTTGCTATATGATGACCAACAATTTGNAAGACAGCCCTACATCCAGTTTATGGG	1342
Qy	1141	ATCTTSATCAATGGGCAAAACCAAAATTTGGAAAAATATCTGGAAAAAGAAAGAACTGTTCCAG	1200
Dp	1343	ATCTTGATCAATGGGCAAAACCAAAATTTGGAAAAATATCTGGAAAAAGAAAGAACTGTTCCAG	1402
Qy	1201	TTTGATGTCCAAAGCTTCGGAAATCTACTGTGACCCAGAACAGAAACAACCGGAGACACGA	1260
Dp	1403	TTTGATGTCCAAAGCTTCGGAAATCTACTGTGACCCAGAACAGAAACAACCGGAGACACGA	1462
Qy	1261	TGTGAGATPCTGGATTT-----TGCCTTAATGGTCCCACTGAATGAGTTCAACT	1311
Dp	1463	TGTGAGATPCTGGATTTAATGAGAGATGGCTTAATGGTCCCACTGAATGAGTTCAACT	1522
Qy	1312	CCAGCTCCCTGTATTTGTCCTCCGGAAAAACCAAGGACTTCAAGGGCCCCAAAGGTGACACCT	1371
Dp	1523	CCAGCTCCCTGTATTTGTCCTCCGGAAAAACCAAGGACTTCAAGGGCCCCAAAGGTGACACCT	1582
Qy	1372	GGACTCCCTGGAAACCTCGGTACCCCTGGACAACCTGGTCAAGATGGTAAAGCCTGGATAT	1431
Dp	1583	GGACTCCCTGGAAACCTCGGTACCCCTGGACAACCTGGTCAAGATGGTAAAGCCTGGATAT	1642
Qy	1432	CAGGGAAATTCAGGACACACCAAGGTGTTCCAGGATCTCCAGGAATACAAAGGACCTCGAGA	1491
Dp	1643	CAGGGAAATTCAGGACACACCAAGGTGTTCCAGGATCTCCAGGATACAAAGGACCTCGAGA	1702
Qy	1492	CTACCAAGTTACAAAGAGAACCAAGGGCGAGATGGTGACAAGGGTGATCGTGGACCTCTCT	1551
Dp	1703	CTACCAAGTTACAAAGAGAACCAAGGGCGAGATGGTGACAAGGGTGATCGTGGACCTCTCT	1762
Qy	1552	GGTTTTCTGGGCTTCATGGCATGCCAGGATCAAGGGTGAAATGGTGCCCAAGAGGAC	1611
Dp	1763	GGTTTTCTGGGCTTCATGGCATGCCAGGATCAAGGGTGAAATGGTGCCCAAGAGGAC	1822
Qy	1612	AAAGGATCACTGGATTTTATGGCAAAAGGGTGCAAAAGGTGCAAAAGGGGAATGCTGGC	1671
Dp	1823	AAAGGATCACTGGATTTTATGGCAAAAGGGTGCAAAAGGTGCAAAAGGGGAATGCTGGC	1882
Qy	1672	TTCCTCGGCTCCCTGGACCTGCTGGAGAACCAAGAGACATGGAAAGGATGGATTAATG	1731
Dp	1883	TTCCTCGGCTCCCTGGACCTGCTGGAGAACCAAGAGACATGGAAAGGATGGATTAATG	1942
Qy	1732	GGTAGTCCCGTTTCAAGGAGAACCAAGGATCCCTGGTCTCCGGGGCAGGATGGAACA	1791
Dp	1943	GGTAGTCCCGTTTCAAGGAGAACCAAGGATCCCTGGTCTCCGGGGCAGGATGGAACA	2002
Qy	1792	CGGGAGAGCCTGGAAATCCAGGATTTCTGGAAACCGAGGATTAATGGGCCAAAGAGGA	1851
Dp	2003	CGGGAGAGCCTGGAAATCCAGGATTTCTGGAAACCGAGGATTAATGGGCCAAAGAGGA	2062
Qy	1852	GAAATTTGGGCTCCAGGACACCAAGAAAAAAGGAGCCCCCAGGGATGCTCTGGTTAATG	1911
Dp	2063	GAAATTTGGGCTCCAGGACACCAAGAAAAAAGGAGCCCCCAGGGATGCTCTGGTTAATG	2122
Qy	1912	GGAGCAATGGCTCCAGGACCGCTGGACACCGGATCTTAAGGGAGCAAGGTTCAA	1971
Dp	2123	GGAGCAATGGCTCCAGGACCGCTGGACACCGGATCTTAAGGGAGCAAGGTTCAA	2182
Qy	1972	CCTGGAAATCAAGGGATGCTCGGGCTTCTGGGCTCAAGGGAGAACAGGACCAACGGT	2031
Dp	2183	CCTGGAAATCAAGGGATGCTCGGGCTTCTGGGCTCAAGGGAGAACAGGACCAACGGT	2242
Qy	2032	TCCCAAGGAGAACCAAGGATAGATGGTTTACCTGGGATTCAGGAAAAAAGGGGACAAA	2091
Dp	2243	TCCCAAGGAGAACCAAGGATAGATGGTTTACCTGGGATTCAGGAAAAAAGGGGACAAA	2302

6

1223 GTTAAAGCGTTGTTGTAAGGCTGGCA CCAATTTGGTCTCTTAGTAACAGAAAGAT 1282  
1081 GTGACTTTGTATATTCATGACCAACAAATTTGAAACCAAGCCCTTACATCCAGTTTAGGG 1340  
1283 GTGACTTTGTATATTCATGACCAACAAATTTGAAACCAAGCCCTTACATCCAGTTTAGGG 1342  
1141 ATCTTGATCAATGGGCAACCCMAATTTGAAATAATTTCTGAAAAGAAAGAACTGTTGAG 1200  
1343 ATCTTGATCAATGGGCAACCCMAATTTGAAATAATTTCTGAAAAGAAAGAACTGTTGAG 1402  
1201 TTTGATGCTCCAAAGTTGGCAATCTACTGTGACCCAGAACAGCAACACCGGAGACAGCA 1260  
1403 TTTGATGCTCCAAAGTTGGCAATCTACTGTGACCCAGAACAGCAACACCGGAGACAGCA 1462  
1261 TGTGAGATCTCGATTT-----TGCCTTAATGGTCCCAAGTGATGTAGTTCAACT 1311  
1463 TGTGAGATCTCGATTTAATGGAGAGTGCCTTAATGGTCCCAAGTGATGTAGTTCAACT 1522  
1312 CCAGCTCCCTGTATTTCTCTCCGGAAACCAAGGACTTCAAGGCCCCCAAGGTTGACCCCT 1371  
1523 CCAGCTCCCTGTATTTCTCTCCGGAAACCAAGGACTTCAAGGCCCCCAAGGTTGACCCCT 1582  
1372 GGAATGCTGGGAACCCCTGCTACCTGGACAACTGGTCAAGATGCTAAGCTCGATAT 1431  
1583 GGAATGCTGGGAACCCCTGCTACCTGGACAACTGGTCAAGATGCTAAGCTCGATAT 1642  
1432 CAGGGAATTCAGGAGACACAGGTTTCCAGGATCTCCAGGAATACAGGAGCTCGAGGA 1491  
1643 CAGGGAATTCAGGAGACACAGGTTTCCAGGATCTCCAGGAATACAGGAGCTCGAGGA 1702  
1492 CTACAGGTTTACAAAGAGAACCAAGGCGAGATGTTGACAAAGGTTGATCTGTTGATCTCT 1551  
1703 CTACAGGTTTACAAAGAGAACCAAGGCGAGATGTTGACAAAGGTTGATCTGTTGATCTCT 1762  
1552 GGTGTTCTGGGCTTCATGGCATGCCAGGATCAAGGTTGAAATGGTGGCCCAAGAGAC 1611  
1763 GGTGTTCTGGGCTTCATGGCATGCCAGGATCAAGGTTGAAATGGTGGCCCAAGAGAC 1822  
1612 AAAGGATCACTGGATTTTATGCAAAAGGGTGCACAAAGGTGCAAAAGGGTAAATCCTGGC 1671  
1823 AAAGGATCACTGGATTTTATGCAAAAGGGTGCACAAAGGTGCAAAAGGGTAAATCCTGGC 1882  
1672 TTCCCTGGGCTTCCTGCACTCTGAGAACACCAAGGAGACATGGAAGATGGATTAATG 1731  
1883 TTCCCTGGGCTTCCTGCACTCTGAGAACACCAAGGAGACATGGAAGATGGATTAATG 1942  
1732 GGTAGTCCCGGTTTCAAGGAGAACGAGGATCCCTGGTCTCCGGGCGAGGATGGAACA 1791  
1943 GGTAGTCCCGGTTTCAAGGAGAACGAGGATCCCTGGTCTCCGGGCGAGGATGGAACA 2002  
1792 CGGGAGAGCTCGAATCCAGGATTTCTGGAACCGAGGATTAATGGGCCCAAGGGA 1851  
2003 CGGGAGAGCTCGAATCCAGGATTTCTGGAACCGAGGATTAATGGGCCCAAGGGA 2062  
1852 GAATTTGGGCTTCAGACAGCAAGGAAAGGAGGAGCCCAAGGATGCTGTTTAATG 1911  
2063 GAATTTGGGCTTCAGACAGCAAGGAAAGGAGGAGCCCAAGGATGCTGTTTAATG 2122  
1912 GGAAGCAATGGCTCACAGGCCACCTGGAAACCGGGATCTAAGGAGAAAGCAAGTGA 1971  
2123 GGAAGCAATGGCTCACAGGCCACCTGGAAACCGGGATCTAAGGAGAAAGCAAGTGA 2182  
1972 CCTGGAATTCAGGAGTCCCTGGGCTTCTGGGCTCAAGGAGAACCAAGGAGCAACGGT 2031  
2183 CCTGGAATTCAGGAGTCCCTGGGCTTCTGGGCTTCAAGGAGAACCAAGGAGCAACGGT 2242  
2032 TCCCAGGAGAACCAAGGATCATGGGTTTACCGGGATTCAGGAGAAAGGGGAGCAAA 2091  
2243 TCCCAGGAGAACCAAGGATCATGGGTTTACCGGGATTCAGGAGAAAGGGGAGCAAA 2302  
2092 GGAATTCAGGAGTAAAGGATTTTCAGGGTCAAGGAGGAGAAATGGAAGACAGGCAATT 2151

6

2303 GGAATCAAGGTGAAAAAGGTATCTCAGGCTCAAAAGGGAGAAAATGGAAGACAGGAATT 2362  
2152 CCAGGCAACAGGGAATTCAGAGCCATCATGCTGTCACAAAGGAGAGAGAGGTGAAAGGGA 2211  
2363 CCAGGCAACAGGGAATTCAGAGCCATCATGCTGTCACAAAGGAGAGAGAGGTGAAAGGGA 2422  
2212 GAACCTGGTGTCCGAGGTCCATTTGGATCAAAAGGAGAAATCTGGGTGGATGGCTTGAATG 2271  
2423 GAACCTGGTGTCCGAGGTGCCATTTGGATCAAAAGGAGAAATCTGGGTGGATGGCTTGAATG 2482  
2272 GGGCCCGCAGGTCCTTAAGGGGCAACCTGGGGATCCAGGTCTCTCAGGGACCCCGAGGTTG 2331  
2483 GGGCCCGCAGGTCCTTAAGGGGCAACCTGGGGATCCAGGTCTCTCAGGGACCCCGAGGTTG 2542  
2332 GATGGAGAGCCCGGAGAGAGATTTTCAGAACAAATTTATTCAGAACAAATTTTCAGACATGTA 2391  
2543 GATGGAGAGCCCGGAGAGAGATTTTCAGAACAAATTTATTCAGAACAAATTTTCAGACATGTA 2602  
2392 ATAAAGAGCCAGCTACCACTACCTTACTTCAGAGTGGAAAGAAATAGAAAATTTGTGATCTGC 2451  
2603 ATAAAGAGCCAGCTACCACTACCTTACTTCAGAGTGGAAAGAAATAGAAAATTTGTGATCTGC 2662  
2452 CTGTCCCAACATGGCTCCCGGGTATTTCTGGGCACTCTGGTCCGATAGGCCAGGAGGT 2511  
2663 CTGTCCCAACATGGCTCCCGGGTATTTCTGGGCACTCTGGTCCGATAGGCCAGGAGGT 2722  
2512 CCAGAGGATTAACCTGGTTCAGAGAGAGATGGTGTCTCTGATTTAGTGGGTGTCCT 2571  
2723 CCAGAGGATTAACCTGGTTCAGAGAGAGATGGTGTCTCTGATTTAGTGGGTGTCCT 2782  
2572 GGAATTCAGGTTCTCAGAGGATTTAAAGGCTTACAGAGAGAAATGGGAAAAGGAGC 2631  
2783 GGAATTCAGGTTCTCAGAGGATTTAAAGGCTTACAGAGAGAAATGGGAAAAGGAGC 2842  
2632 CAGGTTTGGGTATCTGGAGAACAGGTCCTCTGTCCTCCAGGTCCTCAGAGGAGCCCT 2691  
2843 CAGGTTTGGGTATCTGGAGAACAGGTCCTCTGTCCTCCAGGTCCTCAGAGGAGCCCT 2902  
2692 CTTGGAATTAAGCAAGAGAGTCTCTCAGAGAGACCCAGTCTCTCTGCAAGAGATGGAGAC 2751  
2903 CTTGGAATTAAGCAAGAGAGTCTCTCAGAGAGACCCAGTCTCTCTGCAAGAGATGGAGAC 2962  
2752 CATGGAACCTCGAATTCAGAGGCAACAGGAGCCCGGAGGTCCTCTGCAAGAGATGGAGAC 2811  
2963 CATGGAACCTCGAATTCAGAGGCAACAGGAGCCCGGAGGTCCTCTGCAAGAGATGGAGAC 3022  
2812 TGTGTTAGTGTAAATTCAGAGAGAGATCCGTTTCAGAAAAGGACCAAACTATTAG 2865  
3023 TGTGTTAGTGTAAATTCAGAGAGAGATCCGTTTCAGAAAAGGACCAAACTATTAG 3076

6

RESULT 3

AX086127 LOCUS 4160 bp DNA linear PAT 09-MAR-2001  
DEFINITION Sequence 79 from Patent WO0112659.  
ACCESSION AX086127  
VERSION AX086127.1 GI:13275878  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1  
AUTHORS Wiemann, S.  
TITLE Human dna sequences  
JOURNAL Patent: WO 0112659-A 79 22-FEB-2001;  
German Human Genome Project (DE)  
FEATURES  
source 1. .4160  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
BASE COUNT 1271 a 795 c 1026 g 1068 t

6



Qy 2701 AGCAAGAGGTCCTCCAGGAGACCCAGGCTCTCCCTGGCAAGATGGAGACCATGGAATA 2760  
Db 2863 AGCAAGAGGTCCTCCAGGAGACCCAGGCTCTCCCTGGCAAGATGGAGACCATGGAATA 2922  
Qy 2761 CTTGGATCCAGGGCAACAGGCCCCCAGGATCTCGACCCATCATTGTTTGT 2820  
Db 2923 CTTGGATCCAGGGCAACAGGCCCCCAGGATCTCGACCCATCATTGTTTGT 2982  
Qy 2821 GTAATGCCAGAGAGATCCGTTTCAGAAAAGGACCAAACTATTAG 2865  
Db 2983 GTAATGCCAGAGAGATCCGTTTCAGAAAAGGACCAAACTATTAG 3027

RESULT 2  
AF414088 4160 bp mRNA linear PRI 16-NOV-2001  
DEFINITION Homo sapiens collagen XXI (COL21A1) mRNA, complete cds.  
ACCESSION AF414088  
VERSION AF414088.1 GI:15593269

SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 4160)  
Fitzgerald, J. and Bateman, J. F.  
A new FACIT of the collagen family. COL21A1  
FEBS Lett. 505 (2), 275-280 (2001)

2:450665  
MEDLINE 11566190  
PUBMED  
REFERENCE 2 (bases 1 to 4160)  
Fitzgerald, J. and Bateman, J. F.  
Direct Submission  
AUTHORS  
TITLE  
Submitted (23-AUG-2001) Cell and Matrix Biology, Murdoch Children's  
Research Institute, Flemington Rd, Parkville, Vic 3052, Australia

Location/Qualifiers  
1. .4160  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/map="6p12-11"  
1. .4160  
/gene="COL21A1"  
203..3076  
/gene="COL21A1"  
/note="FACIT collagen; similar to protein encoded by  
GenBank Accession Number ALI36624"

/codon\_start=1  
/product="collagen XXI"  
/protein\_id="AA02227.1"  
/db\_xref="GI:15593270"  
/translations="MAHYITFLCMLVLLONSVALEDEVRSSRTAPTDLVFLDGL  
SYSGPENFEIVKMLVNTKNFDIGPKFIQVGVQSDYVPLEIPLSGYSGEHLTA  
AVESILYLGNTVTKGAIOFALDYLFKSSRELTIAVLTGKSDQVKAQAARD  
SKITLPAIGVSTEDAELEATANKSSYFVVEDYIAISKIREVMKQKCEVCP  
TRIPVARDERGDIILGLDVKVKKRQLSPKXINGFEVTSKVLSELSTNVFPE  
LPYSYFVTSRFEKVKIIDLRLTIDRQPIAVTLNGLVKKILLFTTSVINGSVV  
TFANQVTKLFDKGMHRIQLVTEQDVTLYIDQOIENKPLHPLVGLINGTGIGYK  
SGKEETVDFQDLKRIYCDPEONNRETACEIPFNCECLNGSPDVGSTPAPICPPGK  
PGLQPKGDKPLGNPYPGQDGKPGYQGIAGTGVGSGPGIQAGAPGVKGP  
GRGDGDKRLPGPGLHGMPSKSGPMKAGDKSGPYGKGAKEGNAGPGLGP  
PAGEPGRHGDGLMGSPGKSGAGSPAGQDGTGRGPIPGFPNGLMQKGEIGP  
PQQGKAPQPMGLMNSGSPQPTPGSKSGKEPGIQGMPGASGLKGFPGATGSP  
GEPYGMELPGIQGMDKGNQKEKIQGOKENGRIQPGQGIQGHCAKGERGK  
EPGVRAIGSKSGVDGLMGAPGPKGPDGPGPLDGLKGRBESEQFIHQVCT  
DIVRALPVLQSGRIRNCDHCLSHSGSPGIPGPGPIGPEGRPLGLPGRDGVPL  
VGVPRGPRVRLKGLPGRNKGSGQGFVPGSGPGRGPGPGPISKEGPPGDPGL  
PGKGDHGPGLIQGPPGPICDPSLCFSVIARRDPRKPNY"

BASE COUNT 1270 a 796 c 1026 g 1068 t  
ORIGIN

Query Match 73.7%; Score 2112; DB 9; Length 4160;  
Best Local Similarity 99.6%; Pred. No. 0;  
Matches 2862; Conservative 0; Mismatches 3; Indels 5; Gaps 1;  
Qy 1 ATGSGTCACCTATTACATTTCTCTGCATGGTTTGGTGTCTGCTTCTTCAGAAATCTGTG 60  
Db 203 ATGSGTCACCTATTACATTTCTCTGCATGGTTTGGTGTCTCTTCAGAAATCTGTG 262  
Qy 61 TTAGCTGAAGATGGGGAAGTAAGATCAAGTTGTCTGCTACTGCTCCGACAGATTTAGTTTC 120  
Db 263 TTAGCTGAAGATGGGGAAGTAAGATCAAGTTGTCTGCTACTGCTCCGACAGATTTAGTTTC 322  
Qy 121 ATCTTAGATGGCTCTTATAGTGTGGCCACAGAAACTTTGAATAGTGAAGATGGCTT 180  
Db 323 ATCTTAGATGGCTCTTATAGTGTGGCCACAGAAACTTTGAATAGTGAAGATGGCTT 382  
Qy 181 GTCATATCACAAAAAACTTTGACATAGGCGGAAGTTTATTCAGTTGGAGTGGTCAA 240  
Db 383 GTCATATCACAAAAAACTTTGACATAGGCGGAAGTTTATTCAGTTGGAGTGGTCAA 442  
Qy 241 TATAGTGACTACCTGTGCTGGAGATTCCTCTCGGAAGCTATGATTCAGGAGAACTTTG 300  
Db 443 TATAGTGACTACCTGTGCTGGAGATTCCTCTCGGAAGCTATGATTCAGGAGAACTTTG 502  
Qy 301 ACGGCAGCAGTGGAAATCCATCTCTACTTAGGAGAAACAAAGACAGAGAGGAGCCATC 360  
Db 503 ACGGCAGCAGTGGAAATCCATCTCTACTTAGGAGAAACAAAGACAGAGAGGAGCCATC 562  
Qy 361 CAGTTTGGCTCGATTACCTTTTCCCAAGTCTCTCAAGATTTCTGACTAAGATAGCAGT 420  
Db 563 CAGTTTGGCTCGATTACCTTTTCCCAAGTCTCTCAAGATTTCTGACTAAGATAGCAGT 622  
Qy 421<sup>6</sup>GTACTTACCGATGGCAATCCCAAGATCAGCTCAGGATGAGCTCAGGAGCAAGAT 480  
Db 623 GTACTTACCGATGGCAATCCCAAGATCAGCTCAGGATGAGCTCAGGAGCAAGAT 682  
Qy 481 AGTAAGATAACATTTATTTGCTATTGCTGTGTTGCTCAGAAACAGAGATGCCAACTTAGA 540  
Db 683 AGTAAGATAACATTTATTTGCTATTGCTGTGTTGCTCAGAAACAGAGATGCCAACTTAGA 742  
Qy 541 GCTATTGCCAAGAGCTCTGCTACTTATGTTTATGTTTATGTTGAGAGACTATATTGCAATA 600  
Db 743 GCTATTGCCAAGAGCTCTGCTACTTATGTTTATGTTTATGTTGAGAGACTATATTGCAATA 802  
Qy 601 TCCAAATTAAGGGAAGTATGATGAAGCAGAAACTTTGTGAAGATCTGTCTCCCAACAGCA 660  
Db 803 TCCAAATTAAGGGAAGTATGATGAAGCAGAAACTTTGTGAAGATCTGTCTCCCAACAGCA 862  
Qy 661 ATTCCAGTGGAGCTCGTGAAGAGGGGATTTGATATCTTTTAGTTAGATGTAAT 720  
Db 863 ATTCCAGTGGAGCTCGTGAAGAGGGGATTTGATATCTTTTAGTTAGATGTAAT 922  
Qy 721 AAAAAGCTTAAGAAAGATACAGCTTTCACAAAAAAGATAAAGGATATGAAGTAAACA 780  
Db 923 AAAAAGCTTAAGAAAGATACAGCTTTCACAAAAAAGATAAAGGATATGAAGTAAACA 982  
Qy 781 TCAAAGTTGATTTATCAGAACTCAAGCAATGTTTCCAGAGAGTCTTCCTCCATCA 840  
Db 983 TCAAAGTTGATTTATCAGAACTCAAGCAATGTTTCCAGAGAGTCTTCCTCCATCA 1042  
Qy 841 TATGATTTGTCTCTACTCAAGATTTTAACTCAAGAAATTTGGGATTTATGGAGATA 900  
Db 1043 TATGATTTGTCTCTACTCAAGATTTTAACTCAAGAAATTTGGGATTTATGGAGATA 1102  
Qy 901 TTAACCTATTGATGAAGGCCCAAAATAGCAGTTACCTTAAATGGTGTGCGACAAAATCTTA 960  
Db 1103 TTAACCTATTGATGAAGGCCCAAAATAGCAGTTACCTTAAATGGTGTGCGACAAAATCTTA 1162  
Qy 961 TTATTTCACACACAGCGTAATTAAGGCTCAGAAAGTGGTTACCTTTGCTACCCCTCAA 1220  
Db 1163<sup>6</sup>TTATTTCACACACAGCGTAATTAAGGCTCAGAAAGTGGTTACCTTTGCTACCCCTCAA 1222  
Qy 1021 GTTAAGACCTTTGTTGATGAAGGCTGGCAACCAAAATTCGTCTCTTAGTAAACAGAACAGAT 1080





C 66	21	0.7	150236	2	AC023292	AC023292 Homo sapi
C 67	21	0.7	157711	9	AC055738	AC055738 Homo sapi
C 68	21	0.7	159160	2	AC024715	AC024715 Homo sapi
C 69	21	0.7	169986	2	AL353589	AL353589 Homo sapi
C 70	21	0.7	172506	2	AC097010	AC097010 Sus scrofa
C 71	21	0.7	175099	2	AC017029	AC017029 Homo sapi
C 72	21	0.7	177476	2	AC114629	AC114629 Mus muscu
C 73	21	0.7	178033	9	AC107025	AC107025 Homo sapi
C 74	21	0.7	178415	2	AC023547	AC023547 Homo sapi
C 75	21	0.7	179523	2	AC026104	AC026104 Homo sapi
C 76	21	0.7	179854	2	AC143471	AC143471 Macaca mu
C 77	21	0.7	183104	9	AL355601	AL355601 Human DNA
C 78	21	0.7	186830	2	AC121539	AC121539 Mus muscu
C 79	21	0.7	188247	9	HS570819	AL442638 Homo sapi
C 80	21	0.7	190394	9	AL157834	AL157834 Human DNA
C 81	21	0.7	191463	9	AC078814	AC078814 Homo sapi
C 82	21	0.7	196127	2	AC132788	AC132788 Rattus no
C 83	21	0.7	203978	2	AC109286	AC109286 Mus muscu
C 84	21	0.7	207945	9	CNS010DS	AL121841 Human chr
C 85	21	0.7	208026	9	AP006292	AP006292 Homo sapi
C 86	21	0.7	213082	2	AC130126	AC130126 Rattus no
C 87	21	0.7	219184	2	AC132830	AC132830 Mus muscu
C 88	21	0.7	219358	2	AC135740	AC135740 Rattus no
C 89	21	0.7	220688	2	AC116463	AC116463 Mus muscu
C 90	21	0.7	234549	2	AC105497	AC105497 Rattus no
C 91	21	0.7	251244	2	AC103149	AC103149 Rattus no
C 92	21	0.7	286957	2	AC131144	AC131144 Rattus no
C 93	20	0.7	393	6	AX507730	AX507730 Sequence
C 94	20	0.7	400	11	GI6945	GI6945 human SIS S
C 95	20	0.7	448	9	AF463542	AF463542 Gorilla g
C 96	20	0.7	548	6	AR225522	AR225522 Sequence
C 97	20	0.7	548	6	AX321643	AX321643 Sequence
C 98	20	0.7	637	11	BV000957	BV000957 S208P6187
C 99	20	0.7	954	6	AR310334	AR310334 Sequence
C 100	20	0.7	954	6	AX097721	AX097721 Sequence

# ALIGNMENTS

RESULT 1	AF330693	4122 bp	mRNA	linear	PR: 01-MAR-2002
LOCUS	AF330693	Homo sapiens alpha 1 chain-like collagen COLA1L precursor (COLA1L)			
DEFINITION	mRNA, complete cds.				
ACCESSION	AF330693.1	GI:17974509			
VERSION					
KEYWORDS					
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	Chou, M.-Y. and Li, H.-C.				
TITLE	Genomic organization and characterization of the human type XX collagen (COL21A1) gene				
JOURNAL	Genomics 79 (3), 395-401 (2002)				
MEDLINE	21853310				
PUBMED	11863369				
REFERENCE	2 (bases 1 to 4122)				
AUTHORS	Chou, M.-Y.				
TITLE	Cloning and identification of a novel human alpha 1 chain-like collagen (COLA1L) gene, a new member of the FACIT family				
JOURNAL	Unpublished				
REFERENCE	3 (bases 1 to 4122)				
AUTHORS	Chou, M.-Y.				
TITLE	Direct Submission				
JOURNAL	Submitted (20-DEC-2000) Biomedical Engineering Center, Industrial Technology Research Institute, 195-85, Sec. 4, Chung Hsing Rd., Chung, Hsinchu 310, Taiwan				
FEATURES	Location/Qualifiers				
source	1..4122				
	/organism="Homo sapiens"				

/mol_type="mRNA"	
/db_xref="taxon:9606"	
/chromosome="6"	
/map="6p12.3-p11.2"	
..4122	
/gene="COLA1L"	
163..3327	
/gene="COLA1L"	
/codon_start=1	
/product="alpha 1 chain-like collagen COLA1L precursor"	
/protein_id="AA150033.1"	
/db_xref="GI:17974513"	
/translation="MAHVTLCVVLVLLNSVLAEDGVRSSCRAPATDLFLDGG	
SYSGPENFEVKKWKNITKFDLGFKFGVGVVQYSDYFVLEIPGSGTDSSEHLTA	
SVESLYLGGNTKGAIOFALDFAKSRFLTKIAVLTGKSQDDVKDAQAAVCP	
SKITLFAIGVSETEDAELRAINKPSTVYFYVEDYIA-SKIREVVKLCEESVGP	
TRIPVAARDGDFDILGLDVKVKKRIQLSPKKIKGVEYTKVDELSTVSGVPEG	
LPPSVFVSTQRFVKKKMLLWRLTIDGRPOIAVTLNGVDKILLFTTTSVNGSQV	
TFANQVVTLDDEGWHQRLAVTQDVTLYIDDQIENKLPVGLILNGQTOIGKY	
SKKETVQFDVKLRIYCDPSONREYACELPGFLNGPSDVSTAPALCPKPKGL	
QPKGDDPLGNPGYPQGGQDGKPGYQGIAGTGVFGSGPGIQARGLPYKKEPGRD	
GDKRGRLPGFPLHMPGSGKMGAKGKSGPGYFKKGAKEKGNAGPPLGPPGAG	
EPGRHGDLMGSPGFKEAGSPQGDGTGTPGPGPNRNGMGKEIGEPGPGQ	
QKKGAPMPGLMGSNGSPGPGTSGKSGKEGPGIQMPGASGLKGPATGSPGP	
GYMGLPGTGGKKGNOGKGIQKQKNGRQIGQGGIQGHGKGRGKGEPE	
VRGATSGESGVDLGMGAPGPGDPPGPGPGLDGKPREFSQFIRVCTDVI	
PAQLPVLLQSGRIKNCOHCLSHQSGPGIPGPGPIGPGRGLPLGPERDGVPLGV	
*PGRPGVRLGKGLPGRNGEKSGSGYGEQGGPPGPPGPPGPGISKEGPPDGLPSK	
DDDHKPGIQGQPPPG:CDPSLCFSVIARNDPFRKGFY"	
163..228	
/gene="COLA1L"	
4072..4077	
/gene="COLA1L"	
BASE COUNT 1280 a 755 c 996 g 1077 t	
ORIGIN	

Query Match:	100.0%;	Score 2865;	DB 9;	Length 4122;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 2865;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	ATGGCTCACTATTACATTTCTCTGGCATGTTTGGTCTGCTCTCTCAGAAATCTGTG	60	
Db	163	ATGGCTCACTAATTACATTTCTCTGCATGTTTGGTCTGCTCTCTCAGAAATCTGTG	222	
Qy	61	TTAGCTGAAGATGGGGAAGTAAAGTCAAGTGTCTACTGCTCGACAGATTAGTTTTC	120	
Db	223	TTAGCTGAAGATGGGGAAGTAAAGTCAAGTGTCTACTGCTCGACAGATTAGTTTTC	282	
Qy	121	ATCTTAGATGGCTCTTATAGTGTGGCCCAAGAACTTTGAATAGTGAAGAAAGTGGCTT	180	
Db	283	ATCTTAGATGGCTCTTATAGTGTGGCCCAAGAACTTTGAATAGTGAAGAAAGTGGCTT	342	
Qy	181	GTCAATATCAGAAAAAATTTGACATAGGGCCGAAGTTTATCAAGTGGAGTGGTCAA	240	
Db	343	GTCAATATCAGAAAAAATTTGACATAGGGCCGAAGTTTATCAAGTGGAGTGGTCAA	402	
Qy	241	TATAGTGAATACCTGTGCTGGAGATTCTCTCGAAGCTATGATCAGGAGAACATTG	300	
Db	403	TATAGTGAATACCTGTGCTGGAGATTCTCTCGAAGCTATGATCAGGAGAACATTG	462	
Qy	301	ACGGCAGCAGTCCGAATCCATCTCTACTTAGAGAGAAACACAAAGACAGGAGCCATC	360	
Db	463	ACGGCAGCAGTCCGAATCCATCTCTACTTAGAGAGAAACACAAAGACAGGAGCCATC	522	
Qy	361	CAGTTTGGCTCGATTACCTTTTGGCAAGTCTCAGCATTTCTSACTAAGATAGCAGTG	420	
Db	523	CAGTTTGGCTCGATTACCTTTTGGCAAGTCTCAGCATTTCTSACTAAGATAGCAGTG	582	
Qy	421	GTACTTACGGATGGCAAAATCCCAAGATGACGTCAAGGATGAGCTCAAGCAAGAGAT	480	
Db	583	GTACTTACGGATGGCAAAATCCCAAGATGACGTCAAGGATGAGCTCAAGCAAGAGAT	642	
Qy	481	AGTAAGATAAATTTATTTGCTATTTGGTCTTTCAGAAACAGAGAGATGCCAATCTAGA	540	

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2003, 19:29:48 ; Search time 10:31 Seconds  
(without alignments)  
11569.053 Million cell updates/sec

Title: US-09-996-611b-5  
Perfect score: 2865  
Sequence: 1 atggctcattactatattacatt.....gaaaggacaaactatttag 2865

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 2888711 seqs, 2045481386 residues

Word size : 0

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database :

GenEmbl:

- 1: gb\_ba:\*
- 2: gb\_htg:\*
- 3: gb\_in:\*
- 4: gb\_om:\*
- 5: gb\_ov:\*
- 6: gb\_pat:\*
- 7: gb\_ph:\*
- 8: gb\_pl:\*
- 9: gb\_pr:\*
- 10: gb\_ro:\*
- 11: gb\_sts:\*
- 12: gb\_sv:\*
- 13: gb\_un:\*
- 14: gb\_vi:\*
- 15: em\_ba:\*
- 16: em\_fun:\*
- 17: em\_hum:\*
- 18: em\_in:\*
- 19: em\_mu:\*
- 20: em\_om:\*
- 21: em\_or:\*
- 22: em\_ov:\*
- 23: em\_pat:\*
- 24: em\_ph:\*
- 25: em\_pl:\*
- 26: em\_ro:\*
- 27: em\_sts:\*
- 28: em\_un:\*
- 29: em\_vi:\*
- 30: em\_htg\_hum:\*
- 31: em\_htg\_inv:\*
- 32: em\_htg\_other:\*
- 33: em\_htg\_mus:\*
- 34: em\_htg\_pln:\*
- 35: em\_htg\_rnd:\*
- 36: em\_htg\_mam:\*
- 37: em\_htg\_vrt:\*
- 38: em\_sy:\*
- 39: em\_htg\_hum:\*
- 40: em\_htg\_mus:\*
- 41: em\_htg\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2865	100.0	4122	9	AF310693	AF310693 Homo sapi
2	2112	73.7	4160	9	AF414088	AF414088 Homo sapi
3	2061	71.9	4160	6	AX085127	AX085127 Sequence
4	2061	71.9	4160	9	HSM8C1597	AX136624 Homo sapi
5	974	34.0	1184	6	AX430410	AX430410 Sequence
6	923	32.2	1169	6	AX430281	AX430281 Sequence
7	890	30.7	2293	9	AK096444	AK096444 Homo sapi
8	749	26.1	1115	6	AX552238	AX552238 Sequence
9	552	19.3	11000	9	AF439327_2	Continuation (3 of
10	552	19.3	126501	9	HS708F5	AX721209 Sequence
11	382	13.3	1414	6	AX721209	AX721209 Sequence
12	252	8.8	447	6	BDC58397	BDC58397 Secrated
13	204	7.1	81696	9	AF438327_3	Continuation (4 of
14	204	7.1	118431	9	HS682J15	ALC34452 Human DNA
15	131	4.6	246	6	BDO53596	BDO53596 Sequence
16	63	2.2	1103	6	AX721212	AX721212 Sequence
17	58	2.0	25483	9	AL450431	AL450431 Human DNA
18	58	2.0	50612	9	AL513530	AL513530 Human DNA
19	24	0.8	128960	9	HSJ919F19	AL109947 Human DNA
20	23	0.8	67884	2	AC090561	ACC90561 Homo sapi
21	23	0.8	132870	8	AC119149	AC119149 Oryza sat
22	23	0.8	155807	9	AF005716	AF005716 Homo sapi
23	23	0.8	166093	9	AC106013	AC106013 Homo sapi
24	23	0.8	188326	9	AC104958	AC104958 Homo sapi
25	23	0.8	304110	8	AE017074	AE017074 Oryza sat
26	22	0.8	2733	10	AY158896	AY158896 Rattus no
27	22	0.8	5551	10	AF272662	AF272662 Rattus no
28	22	0.8	6114	10	GRUFA1V	M76730 Chinese ham
29	22	0.8	7228	10	AF049880	AF049880 Mus muscu
30	22	0.8	51870	2	AC104220	AC104220 Mus muscu
31	22	0.8	52855	2	AC129491	AC129491 Mus muscu
32	22	0.8	111799	9	HSJ850015	AL049637 Human DNA
33	22	0.8	127277	9	AC010175	AC010175 Homo sapi
34	22	0.8	139273	9	AL596275	AL596275 Human DNA
35	22	0.8	140119	10	AL928839	AL928839 Mouse DNA
36	22	0.8	15807	10	AC122459	AC122459 Mus muscu
37	22	0.8	15793	2	AC102109	AC102109 Mus muscu
38	22	0.8	162572	9	AC009533	AC009533 Homo sapi
39	22	0.8	166751	10	AC087066	AC087066 Rattus no
40	22	0.8	169947	2	AC102838	AC102838 Mus muscu
41	22	0.8	176694	2	AC053494	AC053494 Homo sapi
42	22	0.8	177135	10	AL663059	AL663059 Mouse DNA
43	22	0.8	181742	10	AC122397	AC122397 Mus muscu
44	22	0.8	187999	2	AC028928	AC028928 Homo sapi
45	22	0.8	188162	9	CNS05TDR	AL357092 Human chr
46	22	0.8	194477	2	EX248318	EX248318 Danio rer
47	22	0.8	206032	2	AC134799	AC134799 Rattus no
48	22	0.8	207366	10	AL606750	AL606750 Mouse DNA
49	22	0.8	208122	2	AC121207	AC121207 Rattus no
50	22	0.8	210586	2	AC107788	AC107788 Mus muscu
51	22	0.8	219899	10	AL663039	AL663039 Mouse DNA
52	22	0.8	232649	2	AC108267	AC108267 Rattus no
53	22	0.8	241372	2	AC114198	AC114198 Rattus no
54	22	0.8	245306	2	AC123138	AC123138 Rattus no
55	22	0.8	250023	2	AC111499	AC111499 Rattus no
56	22	0.8	254778	2	AC126532	AC126532 Rattus no
57	22	0.8	255369	2	AC095247	AC095247 Rattus no
58	22	0.8	263088	2	AC130872	AC130872 Rattus no
59	21	0.7	6273	10	AC052161	BC052161 Mus muscu
60	21	0.7	7057	10	MUSA1CC	D38162 Mouse mRNA
61	21	0.7	43613	3	AF106575	AF106575 Caenorhab
62	21	0.7	82700	2	AC015635	AC015635 Homo sapi
63	21	0.7	83000	9	AP030355	AP030355 Homo sapi
64	21	0.7	110000	2	AC114905_2	Continuation (3 of
65	21	0.7	119265	9	AL158046	AL158046 Human DNA



**TITLE**  
Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
**JOURNAL**  
Nature 420, 563-573 (2002)  
**REFERENCE**  
6 (bases 1 to 3405)  
**AUTHORS**  
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,  
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,  
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,  
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,  
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,  
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,  
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,  
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,  
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,  
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,  
Takeda, Y., Tanaka, T., Tomaru, A., Toyota, T., Yasunishi, A.,  
Yuramatsu, M., and Hayashizaki, Y.

**TITLE**  
Direct Submission  
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of  
Physical and Chemical Research (RIKEN), Laboratory for Genome  
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
Kanagawa 230-0445, Japan (E-mail: genome-res@gsr.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,  
Fax: 81-45-503-9216)  
CDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in Riken.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Please visit our web site for further details.  
URL: http://genome.gsc.riken.go.jp/  
URL: http://fantom.gsc.riken.go.jp/  
Location/Qualifiers

**FEATURES**  
source  
1..3405  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="FANTOM DB:5930402D07"  
/db\_xref="taxon:10090"  
/clone="5930402D07"  
/tissue\_type="forelimb"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="13 days embryo"  
156..2921  
/note="unnamed protein product; procollagen, type IX,  
alpha 1 (MGDI:88465, GB|NM\_007740, evidence: BLASTN,  
99%, match=3052)  
putative"  
/codon\_start=1  
/protein\_id="BAC27238.1"  
/db\_xref="GI:26126899"  
/translation="MNFWKISVFFCVCSCLGPWVATLKRPRFPANISNGSELCL  
PKIRIQQDLPGFDLISQFIEKASRRITQWVGSTALQVAKLGVNDFRIPRHL  
YPSGLPEEYSLFTFRMTGSLTEKHNIWQIQDSAGREQGVKINGQTKSAFYSKGL  
DGLQTAFLNLSLDFSRHKLMIQVTSATLFTDCIRIESLIPKRGQIDADGFA  
VLGLVNDPOVSVFLEWMLJHCDPLRPRRTCHLPIRTITTSOTIDBRGPGEQGL  
GSPGPGVPGIGIDGDRGKPGPGPGPGPKGPKGPKGPGPGADLGTGPDGSP  
GSPGPGKQGERVPSGRFPGRGIRPGPGPGTGLPGLGRVPGIDPKRPGPG  
PGPPPGTIGFDHDFLPCNSPCPORSOTPGLPKMKHKGAKGEIPGPGHKEE  
GDQELGEVGGPGPGGLRGTIGVDBKKGARGFDGEPGPGIPGAAGDQDQGG  
PPGTGEPGDRGIGGSRGIPGSGKGTGLPQVDRDGIQPMGTGKAGKPGPGD  
VGLQGLPVPGIPCAKVGAKGNTGAPKGLGSSKPGQGGPGGPGVPGRGLP  
GSRGPGVPGSPGIPGKLSVSGFPLGPGPLPGMKGDRGVGEPGPKGEQASG  
BEAGARGDLGMDGPGPKGSVGNFPGRLGPGIRGLPGVEGPGPGPGMGQD  
QKATGLPGIQQPGPAPTDOHIIKQVCMVUVQHFEMALSLKRPDTGASGIPGPGP  
GPGPGPGENGFPQMGIRLPGIKGPGALGLRGPKDLGKXGKGRPGRGPGLGPA  
IGLPDPGPASVYGRNGRDGEGPPGPGVAGIPGVPGPPGPPGPGFCEQASCTLQSQRA  
FSKGPDPK"

**FEATURES**  
polyA\_signal  
3383..3388  
/note="putative"  
polyA\_site  
3405  
/note="putative"  
BASE COUNT  
863 a 819 c 1015 g 708 t

ORIGIN

Query Match 10.1%; Score 289.6; DB 11; Length 3405;  
Best Local Similarity 50.3%; Pred. No. 2e-44;  
Matches 749; Conservative 0; Mismatches 729; Indels 12; Gaps 1;  
Qy 1318 CCCTGTAATTTGCTCCCGGAAACACAGGACTTCAAGCCCAAGGTGACCCCTGGACTG 1377  
Db 1389 CCAATTTCTGCCACACAGGTGCTTGGATATCAGGCTTACAGGCATGAAGGCCAT 1448  
Qy 1378 CTTGGAAACCTTGGTACCTTGGACAACTCGTCAAGATGTAAGCTCGATATCAGGGA 1437  
Db 1449 AAAGGGCGAAAGGAGAAATCGCGAGCAGGAAGACAAGGAGCAAGAGGTGAAGAGGT 1508  
Qy 1438 ATTGAGGAGACACACAGGTGTTCCAGGATCTCAGGAATACAGAGACTCAGGACTACCA 1497  
Db 1509 GACCAGGGGAACTCGGAGAAAGTTGGCGATCAAGGACCTCAGAGCTTCAAGGTCTGAGA 1568  
Qy 1498 GGTTCAGAAAGGAGAACACAGGCGAGATGGTACAGGCGTATCGTGGACTTCTTGGTTT 1557  
Db 1569 GGCATCACTGGCATAGTTGGAGACAAAGGAGGAGGAGGATTTGATGGAGAG 1628  
Qy 1558 CTTGGGCTTCATGGCATGCGAGGATCAAGAGGTGAATGGTGCCTAAGAGAGACAAGGA 1617  
Db 1629 CTTGGACCTTCAGGCGATTTCCAGGTGAGTCAAGGACAGCGAGGCGCTTCCAGGA 1688  
Qy 1618 TCACCTGTAATTTATGGCAAAAGGTTGCAAAAGGTGAAAAGGGAATGCTGGCTTCCCT 1677  
Db 1689 GAACAGGTCTCAGGAGAGACAGAGGATTCAGGTTCCCGAGGAATTCCTGATCCCA 1748  
Qy 1678 GGCCTCCCTGGACCTGCTGGAGAACACAGAGAGACATGAAAGGATGATTAATGGTGTG 1737  
Db 1749 GGGCAAAAAGGAGACACAGGCTTCCAGGTGAGTGGCCGAGATGGAATACCAAGGAATG 1808  
Qy 1738 CCGGTTTCAAGGAGAGACAGGATCCCTGCTCCCGGAGAGATGGAACACAGGGA 1797  
Db 1809 CCGGAGCAAAAGGTGAGAGAGAGAGGAGCTGACCCCCCGGGAGCTGGATTCAGAGGC 1868  
Qy 1798 GAGCTTGAATCCAGGATTTCTTGGAAACCGAGGATTAATGGCCCAAAAGGAGGAAT 1857  
Db 1869 TTACAGGTGTCTCTGGATCTCTGTCAAAAGGTGTGCGGTGAAAAGGGTAAACAG 1928  
Qy 1858 GGCCTCCAGGACAGCAAGGAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1917  
Db 1929 GGTGCTCCAGGAAAGCCCGGTGAGTGGGAAAGTTCAGGGAACACAGGAGAGAGAGAG 1988  
Qy 1918 AATGGCTCACCAGGAG 1977  
Db 1989 CCAGGAG 2048  
Qy 1978 ATTCAAGGAG 2037  
Db 2049 CCAGAGAGGTCTCCAGGAGATACAGGAGAACTGGGATCTGTGGAGAGAGAGAGAGAGAG 2108  
Qy 2038 GGAGAACACAGGATACATGGTGTATCCCGGGAATCAAGGAAAAAGAGAGAGAGAGAGAGAG 2097  
Db 2109 GGCCTGCTGGGAG 2168  
Qy 2098 AAGGTGAAAAAGGTATTCAGGAGTCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2157  
Db 2169 CCGGCTCCCAAGGAGTGAACAGGAGTCCCTCTGTTGAGAGAGAGAGAGAGAGAGAGAGAG 2228  
Qy 2158 CAACAGGAGATTCAGAGGAGATCATGGTCAAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 2217  
Db 2229 GACCTTGGAGATATGGGAGAACCTGCGCCCAAGGAGATCTGGGTAAACCCCGGAGAGAGAG 2288  
Qy 2218 GGTGCTCCAGGAGTGCATGGATCAAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2277  
Db 2289 GGTCTCAGGAGGCTTGAAGAGATCAGAGGCTTCTGAGTGAAGAGAGAGAGAGAGAGAGAG 2348  
Qy 2278 GAGGTCTCAAGGAG 2337  
Db 2349 CTTGAG 2408

Db 972 GAACAAGGTCCTGAGGAGACAGAGGCAITTCAGGTTCCCGAGGAATTCCTCGATCCCA 1031  
Qy 1678 GGCCTCCTCGACCTGCTGGAGAACAGGAGACATGGAAGGATGATTAATGGGTAGT 1737  
Db 1032 GGGCCAAAAGSAGACACGGCTTGCCAGGTGTAGATGGCCGAGATGGAATACAGGAATG 1091  
Qy 1738 CCCGGTTTCAAGSAGAGAGAGATCCCTGGTGTCTCCGGGCGAGGATGGAACACGGGA 1797  
Db 1092 CCCGGGACAAAGAGGTGAAGCAGGGAAGCCTGGACCCCGGGGACCTGGGATTCAGGGC 1151  
Qy 1798 GAGCTCGAATCCAGGATTCCTGGAAACCGAGGATTAATGGGCCAAAGAGGAGAAAT 1857  
Db 1152 TTACGAGTGTCTGATCCCTGGTGTGAAAGGTCTTCCCGTCAAGAGGTTAACAG 1211  
Qy 1858 GGGCTCAGAGACAGAGGAAGAAAGAGAGCCCGAGGATGCTGGTTAATGGGAAGC 1917  
Db 1212 GGTCTCCAGGAAGCCCGGTCACTTGGGAAGTTTCAGGGAACACAGGCCAACAGGGCC 1271  
Qy 1918 AATGGCTCACAGCCAGCTCGGAACACCGGGATCTAAGGGAAGCAAGTGAACCTGA 1977  
Db 1272 CCAGAGAGGTTCGACCTCGGGACCCAGGGCTCTCCAGGCACTAGAGCCCGTAGA 1331  
Qy 1978 ATTCAAGGATGCTCGGGCTTCTGGGCTCAAGGAGAACACAGGACAAAGGTTCCCA 2037  
Db 1332 CCAGAGGGTCTCCAGGCATACACAGGGAATCTGGATCTTGGCAGCCCTGGCTTCT 1391  
Qy 2038 GGAGAACAGGATACATGGGTTTACCGGATTCAGGGAAGGGAAGGACAAAGGAAT 2097  
Db 1392 GCTTGTCTGGCCCTCGGACTCTCTGGATGAAGGAGACAGGGGTGTATTTGGTGA 1451  
Qy 2098 CAAGGTGAAGAGTATTCAGGTCAAAGGGGAGAAATGAAGAGACAGGGAATTCAGG 2157  
Db 1452 CCGGCTCCAGGGTGAACAGGGTCTCTGGTGAAGAGTGAAGCAGCAGCAAGGGT 1511  
Qy 2158 CAACAGGATTCAGGCTCATGGTGCRAAGGAGAGAGATGGAAGGAGCACT 2217  
Db 1512 GACCTTGAGATATGGACACCTGGCCCAAGGATCTGTGGTAAACCCGGGAGCCG 1571  
Qy 2218 GGTCTCGAGTGCATTTGGATCAAGAGGAGAAATCTGGGTGGATGGCTTGTAGGGGCC 2277  
Db 1572 GGTCTGAGGGGCTGAAGGAATCAGAGGCTCTCTGGTGAAGGACCAAGAGGACCA 1631  
Qy 2278 GCAGTCTTAAGGGGCACTGGGGATCAGGTCCTCAGGACCCCGAGTTGGATGGG 2337  
Db 1632 CTTGACCCCGGCATCGAGGAGATCAGGGTCCACTGGCTGCTGATTCAGGCGC 1691  
Qy 2338 AAGCCGGAAGAGATTTTCAGAACAAATTTATTCGACAGTTTGCACAGATGTAATAAG 2397  
Db 1692 CTTGAGGAGAGCCGACCGACGACATCAGCAGGTTTGTATGAGTCTGTGAA 1751  
Qy 2398 GCCAGCTACAGTCTTACTTCAGATGGAAGAAATTAAGAAATGTATGATTCCTGTC 2457  
Db 1752 GAGCATTTTGGGAATGGCAGCAGCTCAAGAGACACAGAGCCT----- 1803  
Qy 2458 CAACATGCTCCCGGATTTCTGGGCACTGCTCCGATAGGCCAGAGGTCCTCAGA 2517  
Db 1804 ----CTGTTCTTCTGGAGGCTTGGCCCTCGGCTCGGCTCGGCTGATTCAGGAGAT 1859  
Qy 2518 GGATTACCTGTTTCCAGGAGAGATGGTGTCTCTGGATTAGTGGGTGCTCCGACGT 2577  
Db 1860 GGTTCCTGTCAGATGGGAATCCGTGTCTCCAGGCATTAAGGTCCTCCCTGGTCT 1919  
Qy 2578 CCAGGTGTCAGAGATTAAGGCTTACAGGAGAGATGGGAAGGAGGAGCCAGG 2637  
Db 1920 CTTGCTTAAGAGACCTTAAGGAGACTTGGGAGAAAGAGAACTGGTCTCCAGGA 1979  
Qy 2638 TTTGGTATCTCGAGAACAGGCTCTCTGGTCCCGCAGGTCAGAGGCTCTCTGGA 2697  
Db 1980 AGAGTCTTAAGGTTTCTGCTGGAGGATAGTCTCTCCAGGTACCCAGGCTCTGACG 2039  
Qy 2698 ATAAGCAAGAGGTCTTCCAGGAGACCCAGGTCCTCTGCGCAAGATGAGACCAATGA 2757  
Db 2040 TATGGGAAAAATGGCGTGAAGGAGCAAGGTCCTCCCGGAGTGGCAGGAATTCCTGGT 2099

Qy 2758 AAACCTTGGAAATCAAGGGCAACCCAGGCCCCCGAGGATCTGCGACCCATC 2807  
Db 2100 GTGCTTGGAGCCCGAGGTCCTCAGGCCCTCTGCGTCTGTSAGCCAGC 2149

RESULT 15  
AK031071 3405 bp mRNA linear HTC 05-DEC-2002  
LOCUS Mus musculus 13 days embryo forelimb cDNA, RIKEN full-length  
DEFINITION enriched library, clone:5930402D07 product:procolagen, type IX,  
alpha 1, full insert sequence.

AK031071 1 GI:26326988  
VERSION AK031071  
KEYWORDS F-3; CAP trapper.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
1 Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning  
Meth. Enzymol. 303, 19-44 (1999).  
99279255  
10349636

REFERENCE  
2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
Itoh, M., Konro, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)  
20499374  
14042159

REFERENCE  
3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,  
Konro, H., Akiyama, J., Nishi, K., Kusunagi, T., Tashiro, H., Itoh, M.,  
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,  
Yamamoto, R., Yatsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Chura, E., Watahiki, M.,  
Yoneda, Y., Shikawa, T., Ozawa, K., Tanaka, T., Matsuda, S., Kawai, J.,  
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system--184-format  
sequencing pipeline with 384 multiplexed sequencer  
Genome Res. 10 (11), 1757-1771 (2000)  
20530913  
11376862

REFERENCE  
4 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,  
Arakawa, T., Harai, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,  
Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yananaka, I.,  
Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R.,  
Kadoya, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,  
Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,  
Kuehl, P., Lewis, S., Matsuo, Y., Nikaide, T., Pesole, G.,  
Quackenbush, J., Schram, L. M., Staubli, F., Suzuki, R., Tomita, M.,  
Wagner, L., Washio, T., Sakai, K., Okido, T., Furusio, M., Aono, H.,  
Badrall, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N.,  
Carninci, P., De Bonaldo, M. F., Brownstein, K. J., Bult, C.,  
Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,  
Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P.,  
Marchionni, L., Mashima, J., Mazzarelli, J., Mombauts, P., Nordone, P.,  
Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,  
Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H.,  
Toyooka, K., Wang, K. H., Weitz, C., Whitaker, C., Wilming, L.,  
Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kotsuki, S.  
and Hayashizaki, Y.  
Functional annotation of a full-length mouse cDNA collection  
Nature 409 (6821), 685-690 (2001)  
21085660  
11217851

REFERENCE  
5 The PANTOM Consortium and the RIKEN Genome Exploration Research  
Group Phase I & II Team.

```

Db 1246 CCAGGAGAGGTTGGACCTCGGGGACCCAGGGGCTTCACGAGCATAGAGGCCCGGTAGGA 1305
Qy 1978 ATTCAAGGAGATCGCTGGGGCTTCTGGGCTCAAGGGAGAACACGAGAACCGGTTCCCCA 2037
Db 1306 CCAGAAAGGTTCTCCAGGCATACCCAGGAACTGGGATCTGTTGGCAGCCCTGGCTTCT 1365
Qy 2038 GGAGAACACAGGATACATGGTTTACCGGGATTCAAGGAAGAGAGGGGCAAGAGAAAT 2037
Db 1366 GGCTTCCTGGGGCCCTGGACTTCTGGAATGAAGGAGACAGGGGTGATTGGTGAA 1425
Qy 2098 CRAAGTGAAAGAGTATTCAAGGTCAAAAGGGAGAGAAATGGAAGACACAGGAATCCAGGG 2157
Db 1426 CCGGTTCCCAAGGGTGAACAGGGTGCTCTGTTGAAGAGTGAAGCAGGAGCAAGGGT 1485
Qy 2158 CAACAGGGAATCAAGGCATCATGTGTGCAAAAGAGAGAGAGTGAAGAGGAGAACCT 2217
Db 1486 GACCTTTGGAGATATGGGCAACCTGGCCCAAGGGATCTGTGGGTAAACCCCGGGAGCGG 1545
Qy 2218 GGTGTCGAGGTGCCATTGGATCAAAAGGAGAAATCTGGGGTGGTGGCTTGATGGGGCC 2277
Db 1546 GGTCTGAGGGGCTGAAGGATCAGAGGCTTCTGGAGTGAAGAGACCAAGAGACCA 1605
Qy 2278 GCAAGTCTTAAGGGCAACCTGGGGATCCAGGTCTCAGGACCCCGGAGGTTTGGATGG 2337
Db 1606 CTTGGACCCCGGGGATGAGGGAGATCAGGGTGCCACTGGGCTGCTGGTATCCAGGGC 1665
Qy 2338 AAGCCCGGAAGAGAGTTTTCAGAACAAATTTATTCGACAGTTCGACAGATCTAATAGA 2397
Db 1666 CTTCCGGGAGAGCGGCGACCGACAGACATCAAGCAGGTTTGATCAGAGTGTGGA 1725
Qy 2398 GCCAGCTACAGTCTTACTTCAGAGTGAAGAAATAGAAATTTGATCATTTGCTGTCC 2457
Db 1726 GAGCAATTTGCGGAATGCGAGCAGCTCAAGAGACCAAGAGAGGAGCTT----- 1777
Qy 2458 CAACATGGCTCCCGGGTATCTCGGCCACCTGTGCTCCATAGGCCAGAGGTCCTCAGA 2517
Db 1778 -----CTGGTCTTCTGGAGGCTTGCCCCCTTGGGCTTCCGGGCCCTCGGAGAGAA 1833
Qy 2518 GGAATTACCTGGTTTCCAGAGAGATGTTGTTCTCTGGATTAGTGGTGTCTCTGGAGT 2577
Db 1834 GGTTCCTCTGTCAGATGGAATCCGTGGTCTCCAGGCAATTAAGGGTCCCCCTGGTCT 1893
Qy 2578 CCAGGTCTCAGAGATTAAGGCTTACAGGAGCAATGCGGAGAAATGCGGAAAGGAGCAAGGG 2637
Db 1894 CTTGGCTTAAGAGACCTTAAGGAGACTTGGGAGAAAGAGAGAGCTGTCTCTCAGA 1953
Qy 2638 TTGGGTATCTCTGAGAACAGGTCTCTCTGTGCTCCCGGAGTCCAGAGGGCCCTCTGGA 2697
Db 1954 AGAGTCTTAAGGTTTGGCTGGAGCATAGGTCTCCAGGTGACCCAGGCCCTGCCAGC 2013
Qy 2698 ATAAGCAAGAGGTTCTCCAGGAGACCCAGGCTCTCTCGGCAAGATGGAGACCATGGA 2757
Db 2014 TATGGGAAATGCGCGTGACGAGAGCAAGGTCCCGGAGTGGCAGGAATTCCTGT 2073
Qy 2758 AAACCTGGAATCAAGGGCAACAGGCCCCCGGAGCATCTGGACCCCATC 2807
Db 2074 GTGCTCGACCCGAGTCTCCAGGCCCTCTCTGGTCTGTGAGCCACC 2123

RESULT 14
BC038017
LOCUS BC038017 3035 bp mRNA linear HTC 23-SEP-2002
DEFINITION Mus musculus, Similar to procollagen, type IX, alpha 1, clone
IMAGE:3592397, mRNA.
ACCESSION BC038017
VERSION BC038017.1 GI:23274264
KEYWORDS HTC.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;
REFERENCE 1 (bases 1 to 3035)
AUTHORS Strausberg,R.

```

TITLE  
JOURNALREMARK  
COMMENT:

Direct Submission  
Submitted (20-SEP-2002) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)

Tissue Procurement: Gilbert Smith, Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: National Institutes of Health Intramural

Sequencing Center (NISC),

Gaithersburg, Maryland.

Web site: <http://www.nisc.nih.gov/>

Contact: [nisc\\_mgc@nsl.nih.gov](mailto:nisc_mgc@nsl.nih.gov)

Akher,N., Ayele,N., Beckstrom-Sternberg,S.M., Benjamin,B.,

Blakesley,R.N., Bouffard,G.G., Green,K., Brinkley,C., Brooks,S.,

Dierich,N.L., Granite,S., Guan,X., Gupta,S., Haghighi,P.,

Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,

Maduro,Q.L., Masiello,C., Maskeri,B., Yastrian,S.D., McCloskey,J.C.,

McDowell,C., Pearson,R., Staatripop,S., Thomas,P.J., Touchman,J.W.,

Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,

Young,A., Zhang,L.-H. and Green,S.D.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAK Plate: 80 Row: c Column: 17

This clone was selected for full length sequencing because it

passed the following selection criteria: Similarity but not

identity to protein

This clone has the following problem: frame shifted.

Location/Qualifiers

..3035

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="FVB/N"

/db\_xref="taxon:10090"

/clone="IMAGE:3592397"

/tissue\_type="Mammary tumor. Metallothionien-TGF alpha

model. 10 month old virgin mouse. Taken by biopsy."

/clone\_lib="NCI CGAP\_Mam1"

/lab\_host="DH10B"

/note="Vector: PCMV-SF0RT6"

BASE COUNT 767 a 716 g 915 g 637 t

## ORIGIN

Query Match 10.2%; Score 291.2; DB 11; Length 3035;

Best Local Similarity 50.3%; Pred. No.1e-44;

Matches 750; Conservative 0; Mismatches 728; Indels 12; Gaps 1;

Qy 1318 CCTGTATTGTTCTCCCGGAAACCCAGGACTTCAAGGCCCAAGGTGACCTGSACTG 1377

Db 672 CCCAATTCCTGCCACCCAGGTCGTCTGTGATATCAGGCTTACCAGGCATGAGGGCCAT 731

Qy 1378 CTGGGAACCTCGGCTACCTGGACACCTGTCAGATGTAAGCCTGGATATCAGGA 1437

Db 732 AAAGGGCCGAAGAGAGAAATCGCGAGCCAGGAACAGACACAGGGTGAAGAGGT 791

Qy 1438 ATTGCAGGACACACAGGTGTTCCAGGATCTCCAGGAATACAGAGGCTCCAGGACTACCA 1497

Db 792 GACCAGGGGAACTGGGAGAAAGTTGGCGATCAAGGACCTCCAGGACCTCAGGCTTGA 851

Qy 1498 GGTTCACAAAGAGAACACAGGCGAGATGGTGACAGGGTGAATCGTGACTTCCTGGTTT 1557

Db 852 GGCATCTCTGGCATAGTTGGAGACAAAGAGAAAGGGTCTCGGGATTTGATGGAGAG 911

Qy 1558 CTTGGCTTCATGGCATGCCAGGATCAAGGGTGAATGGGTGCCAAGAGAGACAAGGA 1617

Db 912 CTTGGACCTCAGGGCATCTCCAGGTGCACTGGTATCAAGGACAGGAGGCCCCCTCCAGGA 971

Qy 1618 TCACCTGGATTTTATGGCAAAAGGGTGCACAAAGGTTGAAGGSGAATGCTGGCTTCC 1677





```

Db      651  CCCAATTCCTGCCCCACCAAGTCTGCTCTGGATATCCAGGCTACCAAGCATCAGGGGCAT 710
QY      1378  CTTGGAACTTCTGGCTACCTGGCAACCTGCTCAAGATGGTAAGCTTGGATATCAGGGA 1437
Db      711  AAAGGGCGAAAGAGAAATCCGCGAGCCAGGAAACAAGGACACAAGGGTGAAGAGGGT 770
QY      1438  ATTGCAGGCACACCAAGGTCTTCCAGGATCTCCAGGAATACAAGGAGCTCAGGAGTACCA 1497
Db      771  GACCAAGGGAACTGGGAAAGTTGGCGATCAAGGACCTCCAGGACCTCAGGCTCTGAGA 830
QY      1498  GGTTCACAAAGGAGAACCAAGGGGAGATGGTGAACAAGGTGATCGTGGACTTCTGGTTT 1557
Db      831  GGCATCACTGGCATAGTTGGAGACAAAGGAGAAAGGGTGTCTCGGGATTTGATGGAGAG 890
QY      1558  CTTGGCTTCATGGGATCCAGGATCAAAAGGTTGAATGGGTGCCAAGGAGACAAAGGA 1617
Db      891  CTTGGACCTCAGGGCATTCAGGTGCAGCTGGTGATCAAGGACACGCGAGGCCCTCCAGGA 950
QY      1618  TCACCTGGATTTTATGGCAAAAGGTTGCAAAAGGTTGAAAGGGGAATGCTGGGTCTCCCT 1677
Db      951  GAAACAGGTCCTGAGGGACAGAGGCAATCAAGGTTCAGGAGGAATTCCTGGATCCCA 1010
QY      1678  GGCCTCCTGGACCTGCTGGAGAAACAGGAGACATGGAAGGATGGATTAATGGGTAGT 1737
Db      1011  GGGCCAAAAGGAGACACGGGCTTGGCAGGTGTAGATGGCCGAGATGGGAATACCAAGGAATG 1070
QY      1738  CCGGTTTCAAGGGAGACAGCATCCCTGCTGCTCGGGGACAGATGGAAACACGGGA 1797
Db      1071  CCGGGACAAAGGGTGAACAGGGAAGCTGACCCCGGGAGCTGGATTTGAGGGC 1130
QY      1798  GAGCTCTGGAATCCAGGATTTCTTGAAACCCAGGATTAATGGGCCAAAGGGAGAAAT 1857
Db      1131  TTACCAAGGTGTCCTGGAATCCTGTCGCAAAAGGTGTGCGGTGAAAGAGGTAAACAG 1190
QY      1858  GGGCTCCAGGACGACAGGAAAGGAGAGCCAGGAGCCAGGAGTCCCTGGTTTATGGGAGC 1917
Db      1191  GGTGCTCCAGGAAGCCCGTCACTGGTGGAACTCAGGGAACACAGGCAACAGGGCCC 1250
QY      1918  AATGGCTCACAGGCGAGCTGGAACACCGGGATCTAAGGGGAAGCAAGGTGAACCTGGA 1977
Db      1251  CCAGGAGAGTTGGACTCGGGGACCCAGGGCTTCCAGGCAGTAGAGGCCCGGTAGGA 1310
QY      1978  ATTCAAGGATGCTGGGCTTCTGGGCTCAAGGAGAAACAGGAGAACCGGTTCCCA 2037
Db      1311  CCAGAGGGTCTCCAGGCATACAGGGAACCTGGGATCTGTGGCAGCCCTGGGCTTCT 1370
QY      2038  GGAGAACCAAGGATACATGGTTTACCCGGATTCAGGAAAGGAGAGAGGTGAAGAGGAGAACCT 2097
Db      1371  GACTTGGATATGGGACACTGCTGGAATCAAGGAGACAGGGGTGTATTTGGTGAA 1430
QY      2098  CAAAGGTGAAGAGTATTCAGGTCAAAGGAGAAATGGAAGACAGGGAATCCAGGG 2157
Db      1431  CCGGGTCCCAAGGGTGAACAGGGTGCCTCTGCTGAAGAGGTGAAGAGGAGGAGGAGG 1490
QY      2158  CAAAGGGAATCAAGGCATCATGCTGCAAAAGGAGAGAGAGGTGAAGAGGAGAACCT 2217
Db      1491  GACTTGGATATGGGACACTGCTGGAATCAAGGAGACAGGGGTGTATTTGGTGAA 1550
QY      2218  GGTGTCCGAGGTGCCATTTGGATCAAAAGGAGAAATCTGGGGTGGATGGCTTATGGGGCCC 2277
Db      1551  GGTCTGAGGGGGCTCAAGGAATCAGAGGGCTTCTCTGAGTGGAGAGGACCAAGAGGACCA 1610
QY      2278  GCAGGTCTTAAAGGGCAACTGGGGATCAGTCTCAGGACCCCGAGTTTGAATGGG 2337
Db      1611  CTTGGAACCCCGGGCATGAGGGAGATCAGGGTGCACCTGGCTCTGGTATCCAGGGC 1670
QY      2338  AAGCCCGGAGAGAGATTTTCAGAAATTTATTCAGCAAGTTTGCACAGATGTATAGA 2397
Db      1671  CTTCCGGGAGAGCGCGGACCGACAGCATCAAGCAGTTTGCATAGAGTCTGTGCA 1730
QY      2398  GCCACGCTACAGTCTTACTTCAAGTGGGAAGAAATAGAAATTTGTATCAITGCTCTGTC 2457

```

```

Db      1731  GAGCATTTTGGGAATGGCAGCAGCCTCAAGACACGACACAGGAGCCT----- 1782
QY      2458  CAAATGCTCTCCCGGTATTTCTGGGGCACTGGTCCGATAGGCCACAGAGGTTCCGAGA 2517
Db      1783  ----CTGCTTTCTGGGAGGCTGGCCCCCTGGGCTCGGGCCCCCTGGGAGAGAT 1838
QY      2518  GSATTAACCTGGTTTCCCAAGGAAGATGGTGTCTCTGGATAGTGGGTGTCTCTGGACGT 2577
Db      1839  GCTTTCCCTGGTCAATGGGATCGTGTCTCTCCAGGCAATTAAGGTCCTCCCTGGTGT 1898
QY      2578  CCAGGTGTCAAGGATTTAAAGGCTTACCAAGGAAGATGSGGAAAGGGAGGCAAGGG 2637
Db      1899  CTTGGCTTAAGGAGACCTAAAGGAGACTTGGAGAGAAAGAGAAAGAGGCTGGTCTCCAGGA 1958
QY      2638  TTTGGGTATCTCGGAGAACAAAGGTCTCTCTGTCCTCCAGGTCACAGGAGGCCCTCTGGGA 2697
Db      1959  AGAGTCTTAAGGTTTGCCTGGAGGATAGTCTCCAGGTCACCGAGGCCCTCTGGCAGC 2018
QY      2698  ATAAGCAAGAGGTCCTCCAGGAGACCCAGGTCTCTCTGCAAGATGAGAGCATGGA 2757
Db      2019  TATGGAAATATGCGCTGACGAGCAAGGTCTCCCGGAGTGGCAGGAATCTCTGGT 2078
QY      2758  AACTCTGGATCCAGGGAACCAAGGCCCGGACCATCTCGGACCAATC 2807
Db      2078  GTGGCTGAGACCCCGAGTCTCTCCAGGCTCTCTGGTCTCTGGAGCCAGC 2128

```

## RESULT 13

```

LOCUS      AK011547
DEFINITION Mus musculus 10 days embryo whole body cDNA, RIKEN full-length
            alpha 1, full insert sequence.
ACCESSION  AK011547
VERSION     AK011547.2 GI:26374667
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
AUTHORS     Carninci, P. and Hayashizaki, Y.
TITLE       High-efficiency full-length cDNA cloning
JOURNAL     Meth. Enzymol. 303, 19-44 (1999)
MEDLINE     99279253
PUBMED      10349636
REFERENCE   2
AUTHORS     Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
            Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE       Normalization and subtraction of cap-trapper-selected cDNAs to
            Prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL     Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE     20499374
PUBMED      11042159
REFERENCE   3
AUTHORS     Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
            Konno, H., Akiyama, J., Nishi, K., Kitesuna, T., Tashiro, H., Itoh, M.,
            Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
            Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
            Fujikawa, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Nishikawa, M.,
            Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
            Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
TITLE       PKEN integrated sequence analysis (KISA) system--384-format
            sequencing pipeline with 384 multipillar sequencer
JOURNAL     Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE     20530913
PUBMED      11076861
REFERENCE   4
AUTHORS     Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,
            Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, S., Fukuda, S.,
            Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamahiki, I.,
            Saito, T., Okazaki, Y., Gojohori, T., Bono, H., Kasukawa, T., Saito, R.,
            Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,

```



ORGANISM Homo sapiens

R. R. Williams, A., Wohldmann, P. and Wilson, R.  
 The Washit-Merck Est. Project  
 R. R. Williams, A., Wohldmann, P. and Wilson, R.  
 The Washit-Merck Est. Project

COMMENT:  
Unpublished  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

TEL: 314 236 1800  
FAX: 314 286 1810

This clone is available royalty-free through LLNL ; contact the

Seq primer: T7

### FEATURES

```
source
2: 327
/organism="Homo sapiens"
```

```

/!OI_type="mRNA"
/db_xref="GDB:3896080"

```

```
/db_xref="taxon:9606"  
/close="IMAGE:277704"
```

```
/sex="male"  
/tissue_type="mi:tiple sclerosis"
```

```
/dev_stage="Age 46"
lab_post="DH10B (amplification res.start)"
```

```
/clone_lib="Soares multiple sclerosis 2Nb:MS?"
/noto="Vloetox attam (pharmacia) with a modified
```

polylinker V\_TYPE: phagemid; Site\_1: Not I; Site\_2: Eco RI

```
primer [5'
AAAAAAAAAAGGCTGGTGTGTTTGATGACGCGC
```

double-stranded cDNA was size selected, ligated to Eco RI

the Not I and Eco RI sites of a modified pT7T3 vector

normalization to a Cot = 5. Library constructed by Bento

lesions from one patient was kindly provided by Dr. Kevin

BASE COUNT	82 a	80 c	100 g	64 t	1 others
ASE COUNT	82 a	80 c	100 g	64 t	1 others

Query Match	10.8%;	Score 308;	DB 14;	Per-gm 327;
Best Local Similarity	99.4%;	Pred. No. 5.5e-48;		

```

Matches 319; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
0 0

```

2427 AAGAAATTTGGATCATTTGGCTGTCCCAACATGGCTCCCCGGGATATTCCTGGGCC 2486

1 AAGAAATTAGAAATTG:GATCATTGCCCTGTCCCAACATGGCTCCCCGGGTATTCCTGGGCC 60

2487 ACC TGG TCC GAT AGG CCC CAG AGG TCC CAG AGG ATT ACC TGG TTG CCG AGA GAG ATG G 2546

51 ACCTGGTCCGATA-GCCCAGAGGGTCCCAGAGGATTACCTGGT-TGCCAGGAAGAGATGG 119

2547 TGTTCTGGATTAGTGGGTGTCCTCGACGTCAGGTGTCAGAGGATTAAAGGCCTACC 2606

120 TGTTCCTGGATTAGTGGGTCTCCCTGGACGTCCAGGTGCAGAGGATTAAGGCTACC 179

2607 AGGAAGAAATCGGGAAAAAGGGAGCCCAAGGTTTGGGTATCTCTGGAGACAAGGTCCTCC 2666

b 180 AGAAGAAATGGGGAAGGAGCCAGGTTTGGGTATCCTGGAGACAAGGTCCTCC 239

2667 TGGTCCCCCAGGTCCAGAGGGCCCTCCTGGAATAAGCAAGAAGGTCTCCAGGAGACCC 2726

b  
240 TGGTCCCCCAGGTCCAGAGGGCCCTCTCTGGAAATAGCAAGAGGTCTCTCAGGAGACCC 299

2727 AGGTCTCCCTGGCAAAGATGG 2747

Db 366 GTCTGACTGCATCCCGGGTGGCCCTGGGCTCGTGGGCAAGAGAGATTGCTGGAAAA 425

QY 2615 ATGGGAAAAGGAGGCAAGGGTTTGGGTATCTCTGAGAACAGGTCTCTCTGGTCCCC 2674

Db 426 ATGGTGCAAGGCAATCAAGGAATGGGTTCTCTGGATCCAGGCCACAGACCGC 485

QY 2675 CAGTCCAGAGGGCCCTCTCGAATAAGCAAGAGGTCTCTCCAGGAGACCCAGGTCTCC 2734

Db 486 CAGTCTCAGAGGTCCCGCGTATGATGAAGGAAGGACGCCCTGGAGAGCGTGGGCAGC 545

QY 2735 CTGGCAAGATGGAGCATCGAAGAACTGGAATCCAGGTCAGGCAACAGGCCCCAGGCA 2794

Db 546 CTGGTAAAGATGGAGATCGTGGCAGTCCCGAATGCCAGGACCATGCGGACCCCTGGAA 605

QY 2795 TCTGCGACCCATCACTATGTTTTAGTGTAATTGGCAGAGAGATCCGTTTCAGAAAAAGGAC 2854

Db 606 TCTGTGACCCATCACTGTTTCAGTGTGATGTCGGAAGAGATCCATTAGAAAGGAC 665

QY 2855 CAAA 2858

Db 666 CAAA 669

RESULT 9  
BG699698  
LOCUS  
DEFINITION 602681559f1 NIH\_MGC\_95 Homo sapiens cDNA clone IMAGE:481419 5',  
mRNA sequence.

ACCESSION BG699698  
VERSION BG699698.1 GI:13968272  
KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 744)  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
NIH-MGC <http://mgs.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: Miklos Balkovits, M.D., Ph.D.  
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki  
Toshiyuki and Piero Carninci (RIKEN)  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LHAM10709 row: e column: 16  
High quality sequence stop: 596.  
Location/Qualifiers

FEATURES  
source  
1..744  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone IMAGE:481419  
/tissue\_type="hippocampus"  
/lab\_host="DH10B"  
/clone\_lib="NIH\_MGC\_95"  
/note="Organ: brain; Vector: pBluescriptR (modified  
pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (gtcagag  
); oligo-dr primed using primer 5'-TTTTTTTTTTTTTNN-3',  
size-selected for average insert size 2.5 kb and  
normalized to ROT 5. This is a primary library enriched  
for full-length clones and constructed using the  
Cap-trapper method (Carninci, in preparation). Library  
constructed by M. Brownstein (NIH/NHGRI, National  
Institutes of Health). Note: this is a NIH\_MGC Library."  
202 a 206 c 177 g 159 t

Query Match

12.0%; Score 344; DB 10; Length 744;

Best Local Similarity 96.4%; Pred. NO. 9.2e-55;  
Matches 374; Conservative 0; Mismatches 10; Indels 4; Gaps 2;

QY 1 ATGGGTCTACTATATTACATTTCTCTGCATGTTTTGGTGTCTCTCTCAGAAATCTCTG 60

Db 193 ATGGGTCTACTATATTACATTTCTCTGCATGTTTTGGTGTCTCTCTCAGAAATCTCTG 252

QY 61 TTAGTCTGAAGATGGGAAAGTCAAGTTCTCTGCTACTGCTCCGACAGATTTAGTTTTC 120

Db 253 TTAGTCTGAAGATGGGAAAGTCAAGTTCTCTGCTACTGCTCCGACAGATTTAGTTTTC 312

QY 121 ATCTAGATGGTCTTTAAGTGTGGCCCAAGAACTTTGAA---ATAGTGAAGAAAGTGG 177

Db 313 ATCTAGATGGTCTTTAAGTGTGGCCCAAGAACTTTGAACTAGTGAAGAAAGTGG 372

QY 178 CTGTGCAATATCACAAAAAATTTTCACATAGGGCCGAAGTTTATTCAGAGTTGGAGTGGT 237

Db 373 CTGTGCAATATCACAAAAAATTTTCACATAGGGCCGAAGTTTATTCAGAGTTGGAGTGGT 432

QY 238 CAATATAGTACTACCTCTGCTGAGATTTCTCTCGAAGCTATGATTAGGAGAAACAT 297

Db 433 CAATATAGTACTACCTCTGCTGAGATTTCTCTCGAAGCTATGATTAGGAGAAACAT 492

QY 298 TTGACGG-CAGCAGTGGAAATCCATCTCTACTAGAGGAAACACAAACAGCGGAAGGC 356

Db 493 TTGACGGCCAGCAGTGGAAATCCATCTCTACTTAGGAGGAAACACAAACAGCGGAAGGC 552

QY 357 CATCCAGTTTGGCTCGATTACCTTTT 384

Db 553 CATCCAGTTTGGCTCGCAATACCTTT 580

RESULT 10  
BG108124  
LOCUS  
DEFINITION 603111461f1 CSEQCHL12 Gallus gallus cDNA clone ChEST58j15 5', mRNA  
sequence.

ACCESSION BG108124  
VERSION BG108124.1 GI:25310967  
KEYWORDS EST.

SOURCE Gallus gallus (chicken)

ORGANISM Gallus gallus

REFERENCE 1 (bases 1 to 760)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus.  
Boardman P.E., Sanz-Ezquerro J., Overton, I.M., Eurt, D.W., Bosch, E.,  
Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.C.,  
A Comprehensive Collection of Chicken CDNAS  
Curr. Biol. 12 (22), 1965-1969 (2002)  
22335534  
PUBMED 12445392

COMMENT Contact: Simon Hubbard  
Department of Biomolecular Sciences  
University of Manchester Institute of Science and Technology (UMIST)

PO Box 88, Manchester, M60 1QD, UK  
Tel: 01612008930  
Fax: 01612360409  
Email: [Simon.Hubbard@umist.ac.uk](mailto:Simon.Hubbard@umist.ac.uk).

Location/Qualifiers  
1..760  
/organism="Gallus gallus"  
/mol\_type="mRNA"  
/strain="White Leghorn, Hisex"  
/db\_xref="taxon:9031"  
/clone="ChEST58j15"  
/dev\_stage="36"  
/lab\_host="DH10B"  
/clone\_lib="CSEQCHL12"  
/note="Organ: heart; Vector: pBluescript II KS(+); Site 1:  
EcoRI; Site 2: NotI; Modification of pBluescript II KS(+)  
[Stratagene] vector to accommodate cDNA produced with the

FEATURES  
source

```

Seq primer: M13RPL reverse primer (ABI).
FEATURES
    source
        1..527
            Location/Qualifiers
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="c868a05"
                /tissue_type="RPE/choroid"
                /dev_stage="Adult"
                /lab_host="EMDH10B"
                /clone_lib="Human Retinal pigment epithelium/choroid cDNA
                (Un-normalized, unamplified): cs"
                /note="Organ: Eye; Vector: pCMVSPORT6; Two different donor
                eyes (75-80 years old) yielded approximately 600 mg of
                dissected RPE/choroid tissue. This in turn yielded 340 ug
                of total RNA and 7 ug of mRNA. A directionally cloned cDNA
                library in the pCMVSPORT6 vector was constructed at Life
                Technologies (Rockville, MD; now part of Invitrogen Corp),
                essentially following the protocols of the SuperScript
                Plasmid System (Invitrogen Corp).
                <http://www.invitrogen.com/>. The library code
                designation was cs. For this library, cDNA inserts were
                cloned into the NotI/MluI sites of the vector. EST
                analysis was performed on the unamplified library at the
                NIH Intramural Sequencing Center (NISC)."
BASE COUNT      145 a      126 g      138 t
ORIGIN
    Query Match      13.2%; Score 377; DB 14; Length 527;
    Best Local Similarity 100.0%; Pred. No. 5.1e-61;
    Matches 377; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1  ATGGCTCACTATATTACATTTCTCTGCAATGGTTTGGTGTCTTCTTCCAGAAATCTGTG 60
Db      151 ATGGCTCACTATATTACATTTCTCTGCAATGGTTTGGTGTCTTCTTCCAGAAATCTGTG 210
QY      61  TTAGCTGAAGTGGGAGTAAAGATCAAGTTGCTGCTACTGCTCCGACAGATTAGTTTTC 120
Db      211 TTAGCTGAAGTGGGAGTAAAGATCAAGTTGCTGCTACTGCTCCGACAGATTAGTTTTC 270
QY      121 ATCTTAGATGGCTCTTATAGTGTGGCCAGAAACCTTTGAAATAGTGAAGAGTGGCTT 180
Db      271 ATCTTAGATGGCTCTTATAGTGTGGCCAGAAACCTTTGAAATAGTGAAGAGTGGCTT 330
QY      181 GTCAATATCAAAAAAATTGTACATAGGCCGCAAGTTTATTCAGTTGGAGTGGTTCAA 240
Db      331 GTCAATATCAAAAAAATTGTACATAGGCCGCAAGTTTATTCAGTTGGAGTGGTTCAA 390
QY      241 TATAGTCACTACCTGTGCTGGAGATTCCTCTCGGAGCTATGATTCAGGAGAACATTG 300
Db      391 TATAGTCACTACCTGTGCTGGAGATTCCTCTCGGAGCTATGATTCAGGAGAACATTG 450
QY      301 ACGGCAGAGTGGAAATCCATCTACTTAGGAGGAAACAACAAAGACAGGAGGCCATC 360
Db      451 ACGGCAGAGTGGAAATCCATCTACTTAGGAGGAAACAACAAAGACAGGAGGCCATC 510
QY      361 CAGTTTCGGCTCGATTA 377
Db      511 CAGTTTCGGCTCGATTA 527

RESULT 8
LOCUS      BU250509
DEFINITION 603404356f1 CSEQCHN35 Gallus gallus cDNA clone CHEST304013 5', mRNA
sequence.
ACCESSION  BU250509
VERSION    BU250509.1 GI:25501587
KEYWORDS  EST.
SOURCE    Gallus gallus (chicken)
ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
a (bases 1 to 670)
Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,
Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
A Comprehensive Collection of Chicken cDNAs
Curr. Biol. 12 (22), 1965-1969 (2002)
22335534
MEDLINE
PUBMED
12445392
Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology (UMIST)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.
Location/Qualifiers
    1..670
        /organism="Gallus gallus"
        /mol_type="mRNA"
        /strain="White Leghorn, Hisex"
        /db_xref="taxon:9031"
        /clone="CHEST304013"
        /dev_stage="36"
        /lab_host="DH10B"
        /clone_lib="CSEQCHN35"
        /note="Organ: trunks; Vector: pBluescript II KS(+);
        Site 1: EcoRI; Site 2: NotI; This normalized library was
        constructed from 1 million independent clones. cDNA
        synthesis was initiated using an oligo(dT) primer, using
        methylated C in the first strand synthesis reaction.
        Following this first strand reaction, double-stranded cDNA
        was blunted, ligated to NotI adapters, digested with EcoRI
        , size-selected, and cloned into the NotI and EcoRI
        compatible sites of a custom modified MCS of the
        pBluescript (KS+) vector. The library was normalized in 2
        rounds using conditions adapted from Soares et al., PNAS
        (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
        (1996): 791, except that a significantly longer
        reannealing hybridization was used."
BASE COUNT      0      172 a      162 c      209 g      127 t
ORIGIN
    Query Match      12.4%; Score 355.2; DB 13; Length 670;
    Best Local Similarity 70.9%; Pred. No. 7e-57;
    Matches 471; Conservative 0; Mismatches 193; Indels 0; Gaps 0;
QY      2195 AGAGAGGTGAAGAGGAGAGACCTGGTCTCGAGGTGCCATTGGATCAAAAGAGAAATCTG 2254
Db      6  AGAAGGTGAGAGGGAGAGACCCAGGCATTGAGGCATCAATGACAGAGGAGATCCG 65
QY      2255 GGGTGGATGGCTTGATGGGGCCCGCAGGTCTTAAGGGGCAACCTGGGGATCCAGGTCCTC 2314
Db      66  GTATCCAGGGTCTGGTTGGACCTCTGTGTGACAGGTGAGCTGAGCGTGGAGAGAGAGGGCCCC 125
QY      2315 AGGACCCCCAGGTTTGGATGGGAGCCCGGAGAGAGTTTTCAGACAACTTTATTTCAC 2374
Db      126  CAGGACCCACCTGGATCAGATGGAACCTGCAGAGAAATTTTCAGAGAAATTTATTTCGAC 185
QY      2375 AAGTTTGCACAGATGTAAATAGAGCCAGCTACAGTCTTACTTTCAGAGTGGAGAAATTA 2434
Db      186  AGGTGTGTTGATGTACTAGAACCCAGTTACTGTGATCTCTTACAGATGGAGGCTAC 245
QY      2435 GAAATTGTGATCATTTGCTGTCCCAACATGGTCTCCCGGGTATTCTCTGGGCCACCTGGTC 2494
Db      246  AAAAAGTCAACCACTGTCAGTCCAGAGCGCTTCCCGAGACTTCCAGGACCAACAGGTC 305
QY      2495 CGATAGGCCAGAGGGTCCAGAGGATTACCTGGTTTCCAGAGAGAGATGGTGTTCCTG 2554
Db      306  CAAGAGGCCAGAGAGGTTCCAGAGGGTTTCTCTGGTTTCCAGGGAAATGACGGTGTCCAG 365
QY      2555 GATTAGTGGGTGTCCCTGGGACGTCCTCAGGTGTTCAGAGGATTAAGAGGCTACCAAGAGAA 2614

```

```

Db 121 GAAACTGTCOAATTTGAAGTCCAAAGCTACGAATCTACTGTGACCCAGAGCAGAAAC 180
QY CCGGAGACAGCATGTGAGATCTCTGGATTT-----TGCCTTAATGTCTCCAGTGAT 1299
Db 181 CCGGAGACAGCGGTGTGAATACCTGGATTTAATGTGAGTGCCTCAACGGTCCCACTGAT 240
QY 1300 GTAGGTTCAACTCCAGCTCCCTGTATTGTTCCTCCGGGAAACACAGGACTTCAAGGCCCC 1359
Db 241 GTAGGTTCAACCGCAGCTCCCTGCATTGTGCTCTCCAGMAAACCCAGGACTACAAAGGCC 300
QY 1360 AAGGTCACCTGGACTGCTGGGAACCTGGCTACCTCGACAACCTGCTCAAGATGCT 1419
Db 301 AAGGTCACCCCGACAGCTGCGAATCTCTGCTACCCCGACGGCTGCTCAAGATGCT 360
QY 1420 AAGCTCGATATCAGGGAATTTGAGGACACCCAGGTGTTCCAGGATCTCCAGGAATACAA 1479
Db 361 AAGCTCGATATCAGGGAATAGCAGGATCACAGGTGTTCCAGGATCCCGGAATACAA 420
QY 1480 GGAGCTGGAGACTACAGGTTACAAAGGAGAACCCAGGCGGAGATGTCACAGGATGAT 1539
Db 421 GGAGCACAAAGGACTGCCAGGTTACAAAGGAGAACCTGGGAGAGATGGAGAAAGGAGTAC 480
QY 1540 CGTGGACTTCTCTGCTTTCTCGGCTTCATGGCATGCCAGATCAAAAGGCT 1590
Db 481 CGTGGACTTCTCTGCTTTCTCGGCTTCATGGCATGCCAGATCGAAGGCT 531

RESULT 6
LOCUS BE236040 518 bp mRNA linear EST 10-JUL-2000
DEFINITION 143702 MARC lPIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION BE236040
VERSION BE236040.1 GI:9020758
KEYWORDS EST.
SOURCE Sus scrofa (pig)
ORGANISM Sus scrofa
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
Fahrenkrug,S.C., Smith,T.P.L., Freking,B.A., Cho,J., White,J.,
Vallet,J., Wise,T., Rohrer,G.A., Pertea,G., Sultana,R., Quackenbush
J. and Keele,J.W.
TITLE Porcine gene discovery by normalized cDNA-library sequencing and
EST cluster assembly
JOURNAL Mamm. Genome 13 (8), 475-478 (2002)
MEDLINE 22213789
PUBMED 12226715
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 186, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCTCCAGTCACGACG
Plate: 85 row: G column: 4
Seq primer: ATTTAGTGACACTATAG.
FEATURES
source Location/Qualifiers
1..518
/organism="Sus scrofa"
/mol_type="mRNA"
/db_xref="taxon:9823"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="MARC lPIG"
/note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
Library made from pooled tissue from day 11, 13, 15, 20,
and 30 embryos."
BASE COUNT 157 a 105 c 175 g 81 t

ORIGIN
Query Match 14.7%; Score 420.6; DB 10; Length 518;
Best Local Similarity 88.5%; Pred. No. 2.9e-69;
Matches 456; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 1585 AAGGTGAAATGGTGCCAAAGAGACAAAGGATCACCTGATTTTATGCAAAAGGGT 1644
Db 4 AAGGTGAAATGGTGCCAAAGAGAGATAAAGGATCACCTGATTTTATGCAAAAGGGT 63
QY 1645 GCATAAGGTGAAAGGGGATGCTGGCTTCCTCGGCTCCCTGACCTGCTGGAGAACCA 1704
Db 64 GCGAAGGTGAAAGGGGAGACTGTTGGTTTCTCGGCTTCCTGACGTGCTGGAGAACCA 123
QY 1705 GGAAGACATGGAAGGATGGATTAAATGGTAGTCCCGTTTCAAGGGAGAGCAGGATCC 1764
Db 124 GGAAGACATGGAAGGATGGATTAAATGGTAGTCCCGTTTCAAGGGAGAGCAGGACT 183
QY 1765 CCGTGTGCTCCGGGCGAGGATGGAACACGGGAGAGCTCGAATCCAGGATTTCTCTGGA 1824
Db 184 CCAGAGCCCGCAGGCGAGGATGGATTGCGGGGAGAGCTGGAAATCCAGGATTTCTCTGGA 243
QY 1825 AACCGAGGATTAATGGGCGCAAAAGGAGAAATTTGGGCTCCAGCACAGCAAGGAAAAA 1884
Db 244 AACCGAGGATTAATGGGCGCAAAAGGAGAAATTTGGGCTCCAGCACAGCAAGGAAAAA 303
QY 1885 GGAGCCCGCAGGATGCTGCTGTTTAAATGGGAAGCAATGGCTCACAGGCCAGCCTGGAACA 1944
Db 304 GGAGCCCGCAGGATGCTGCTGTTTAAATGGGAAGCAATGGCTCACAGGCCAGCCTGGAACA 363
QY 1945 CCGGATCTTAAGGAGAACCAAGGTGAACCTGGAAATTAAGGGATGCTCGGGGCTTCTGGG 2004
Db 364 CCAGGACCTTAAGGAGAAATTAAGGTGAACCTGGAAATTAAGGGATGCTCGGGGCTTCTGGG 423
QY 2005 CTCAGGCGAGAACCCAGGAGCAACGGGTTTCCCGCAGAGAACCCAGGATACATCGGTTTACCC 2064
Db 424 CTCAGGCGAGAACCCAGGAGCAACGGGTTTCCCGCAGAGAACCCAGGATACATCGGTTTACCC 483
QY 2065 GCGATTCAGGAGAAAGGGGAGACAAAGGAATCA 2099
Db 484 GCGATTCAGGAGAAAGGGGAGACAAAGGAATCA 518

RESULT 7
CA395709 527 bp mRNA linear EST 06-NOV-2002
LOCUS cs68a05.Y1 Human Retinal pigment epithelium/choroid cDNA
DEFINITION (Un-normalized, unamplified): cs Homo sapiens cDNA clone cs68a05
5', mRNA sequence.
ACCESSION CA395709
VERSION CA395709.1 GI:24731429
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Wistow,G., Bernstein,S.L., Wyatt,M.K., Farris,R.N., Behal,A.,
Touchman,J.W., Bouffard,G., Smith,D. and Peterson,K.
Expressed sequence tag analysis of human RPE/choroid for the
NEIBank Project: Over 6000 non-redundant transcripts, novel genes
and splice variants
Mol. Vis. 8 (4), 205-220 (2002)
JOURNAL 22103460
MEDLINE 12107410
PUBMED 12107410
COMMENT Contact: Wistow G
Section on Molecular Structure and Function
National Eye Institute
6/331, NIH, Bethesda, MD 20892-2740, USA
Tel: 301 402 3452
Fax: 301 496 0078
Email: graeme@helix.nih.gov
Plate: 68 row: a column: 05

```

```

RESULT 4
BX431407      909 bp      mRNA      linear      EST 15-MAY-2003
LOCUS      BX431407 Homo sapiens PLACENTA Homo sapiens cDNA clone CS0DE009YK07
DEFINITION      S-PRIME, mRNA sequence.
ACCESSION      BX431407
VERSION      BX431407.1 GI:30773057
KEYWORDS      EST.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 909)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished
Contact: Genoscope
Genoscope, Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 3377.r for
more information about this cluster, see
http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0BAG0132B02_CS01162_1&cluster=3377.r.
Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0BAG0132B02_CS01162_1.
FEATURES
Location/Qualifiers
1..909
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DE009YK07"
/tissue_type="PLACENTA"
/clone_lib="Homo sapiens PLACENTA"
/notes="Vector: pCMVSPORT6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched,
double-strand cDNA was digested with Not I and cloned into
the Not I and EcoRV sites of the pCMVSPORT 6 vector.
Library was not normalized."
BASE COUNT      251 a 201 c 205 g 249 t 3 others
ORIGIN
Query Match      15.2%; Score 436.8; DB 13; Length 909;
Best Local Similarity 99.3%; Pred. No. 2.7e-72;
Matches 438; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY      2425      GGAGGAATTAGAAATTGTGATCATTCCTGCTCCACATGGCTCCCGGGTATTCCTGGG 2484
Db      30      GNAGAAATTAGAAATTGTGATCATTCCTGCTCCACATGGCTCCCGGGTATTCCTGGG 89
QY      2485      CCACCTGGTCCGATAGGCCAGAGGGTCCAGAGGATTACCTGGTTTCAGGAAGAGAT 2544
Db      90      CCACCTGGTCCGATAGGCCAGAGGGTCCAGAGGATTACCTGGTTTCAGGAAGAGAT 149
QY      2545      GGTGTTCTCGATTAGTGGTGTCCCTGGACGTCCTCAGGTGTTCAGAGGATTAAGGCCTA 2604
Db      150      GGTGTTCTCGATTAGTGGTGTCCCTGGACGTCCTCAGGTGTTCAGAGGATTAAGGCCTA 209
QY      2605      CCAGGAAGAAATGGGGAAAGAGGAGCCAGGGTTTGGGTATCCTCGAGAACAGGTCTCT 2664
Db      210      CCAGGAAGAAATGGGGAAAGAGGAGCCAGGGTTTGGGTATCCTCGAGAACAGGTCTCT 269
QY      2665      CTTGGTCCCACAGTCCAGAGGCCCTCTCTGGAATTAAGCAAGAGGTCCTCCAGGAGAC 2724
Db      270      CTTGGTCCCACAGTCCAGAGGCCCTCTCTGGAATTAAGCAAGAGGTCCTCCAGGAGAC 329
QY      2725      CAGGTTCTCCTGGCAAGATGGAGACCATGGAACCTCGAATCCAGGGCAACAGGC 2784
Db      330      CAGGTTCTCCTGGCAAGATGGAGACCATGGAACCTCGAATCCAGGGCAACAGGC 389
QY      2785      CCCCAGGCATCTGCGACCCATCACTATGTTTTAGTGTAAATTCGCCAGAGAGATCCGTTTC 2844

```

```

Db      390      CCCCAGGCATCTGCGACCCATCACTATGTTTTAGTGTAATTCGCAAGAGATCCGTTTC 449
QY      2845      AGAAAGGACCAACTATTAG 2865
Db      450      AGAAAGGACCAACTATTAG 470

RESULT 5
BE899929
LOCUS      BE899929      531 bp      mRNA      linear      EST 25-APR-2001
DEFINITION      190456 MARC 4BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION      BE899929
VERSION      BE899929.1 GI:10387679
KEYWORDS      EST.
SOURCE      Bos taurus (ccw)
ORGANISM      Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 531)
Smith, T.P.L., Grosse, M.M., Freking, B.A., Roberts, A.J., Stone, R.T.,
Casas, E., Wray, J.E., White, J., Cho, J., Fahrkekrug, S.C., Bennett,
G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chiklo-McKown, C.G.,
Perte, G., Holt, I., Karanycheva, S., Liang, F., Quackenbush, J. and
Keele, J.W.
Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
Genome Res. 11 (4), 626-630 (2001)
21180013
1282978
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: Smith@mail.marc.usda.gov
Single pass sequencing. Bases called and alt. trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACAT
BACKWARD: GTTTCAGTCAGCAGC
Plate: 85 row: E column: 7
Seq primer: ATTAGGTGACACTATAG.
Location/Qualifiers
1..531
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="MARC 4BOV"
/notes="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;
Library made from pooled tissue from day 20 and day 40
embryos."
BASE COUNT      161 a 120 c 146 g 104 t
ORIGIN
Query Match      14.7%; Score 421.4; DB 10; Length 531;
Best Local Similarity 88.7%; Pred. No. 2.1e-69;
Matches 471; Conservative 0; Mismatches 51; Indels 9; Gaps 1;
QY      1569      ACAGAACAGATGTGACTTTGTATATTGATGACCAACAAATTTGAAACCAAGCCCTTACAT 1128
Db      1      ACGGAACAGATGTGACTTCGTATATTGATGATCAACAAATTTGAAACCAAGCCCTTACAT 60
QY      1129      CCAGTTTAGGATCTTGATCAATGGCAACCAATTTGAAATATTCTTGGAAAGAA 1188
Db      61      CCAGTTTAGGATCTTGATCAATGGCAACCAATTTGAAATATTCTTGGAAAGAA 120
QY      1189      GAAACTGTTTCAAGTTTCAAGTTTCAAGTTTCAAGTTTCAAGTTTCAAGTTTCAAGTTT 1248

```



```

QY 1936 CTTGGACACCGGGATCTAAGGGAAGCAAGGTGAACCTGGATTCAGAGGATGCTGGG 1995
Db 428 CTTGGACACCGGGATCTAAGGGAAGCAAGGTGAACCTGGATTCAGAGGATGCTGGG 487
QY 1996 GTTCTGGGCTAAGGGAGAACAGAGCAACCGGTTCCAGGAGAACAGGATACATG 2055
Db 488 GTTCTGGGCTAAGGGAGAACAGGAGCAACCGGTTCCAGGAGAACAGGATACATG 547
QY 2056 GTTTACCCGGGATTCAGGCAAAAAGGGGCAAAAGGAATCAAGGTGAAAAAGTATT 2115
Db 548 GTTTACCCGGGATTCAGGCAAAAAGGGGCAAAAGGAATCAAGGTGAAAAAGTATT 607
QY 2116 CAGGGTCAAAAGGGGAAAAATGGAAGACAGGGAAATCCAGGGCAACAGGGAAATCAAGGC 2175
Db 608 CAGGGTCAAAAGGGGAAAAATGGAAGACAGGGAAATCCAGGGCAACAGGGAAATCAAGGC 667
QY 2176 CATCATGGTCAAAAGGAGAGAGAGGTGAAGAGGAGAACCTGGTCTCCAGGTGCCATT 2235
Db 668 CATCATGGTCAAAAGGAGAGAGAGGTGAAGAGGAGAACCTGGTCTCCAGGTGCCATT 727
QY 2236 GGATCAAAAGGAGAAATCT-GGGGTGGATGGCTTGGATGGGGCCCGCAGGTCTTAAAGGG- 2292
Db 728 GGATCAAAAGGAGAAATCTGGGGTGGATGGCTTGGATGGGGCCCGCAGGTCTTAAAGGGG 787
QY 2293 CAACCTGGGATCCAGGTC---CTCAGGAGACCCCGAGTTGGATGGGAAGCC---GGA 2346
Db 788 AACCTGGGGATCCAGGTCTCTCAGGGGACCCCGAGTTGGATGGGAAGCCCGGAA 847
QY 2347 AGAGAGTTTTCAGAAATTTATTCGACAGTTTGCACAGATGTAATAGAGCCAGCTA 2406
Db 848 AGAGAGTTTTCAGAAATTTATTCGACAGTTTGCACAGATGTAATAGAGCCAGCTA 907
QY 2407 CCAGTCTTACTTCAGAGTGAAGAATTAGAAATTTGATCATTTGCCCTGCCAACATGCC 2466
Db 908 CCAGTCTTACTTCAGAGTGAAGAATTAGAAATTTGATCATTTGCCCTGCCAACATGCC 967
QY 2467 TCCCGGGTATCTCGGGCACTCGTTCGATAGGCCACAGGTCCTCCAGAGATT 2522
Db 968 T-CCCGGGTATTCTGGCCA-CTGGTCCGATAGGCCAGAGGGTCCAGAGATTACT 1021

```

RESULT 3

```

BQ893743
LOCUS
DEFINITION
AGENCOURT_8232060 Lupski dorsal root ganglion Homo sapiens cDNA
clone IMAGE:6182832 5', mRNA sequence.
ACCESSION
BQ893743
VERSION
BQ893743.1 GI:22285757
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 929)
NIH-MGC http://mgc.nci.nih.gov/.
AUTHORS
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cga@nih.gov
Tissue Procurement: Dr. James R. Lupski.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13569 row: h column: 01
High quality sequence stop: 693.

```

FEATURES

```

source
1..929
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"

```

```

/db_xref="taxon:9606"
/clone="IMAGE:6182832"
/sex="male"
/tissue_type="dorsal root ganglia"
/dev_stage="adult, 36 yr"
/lab_host="DH10B"
/clone_lib="Lupski dorsal root ganglion"
Note: Site 2: SalI; cDNA made by oligo-dT priming.
Note: Vector: pCMV-SPORT6 (Life Technologies); Site 1:
Directionally cloned using the following adapters:
5'-TCGACACGCGTCCG-3' and
5'-GACATAGTTAGTCAGGAGCGCCGCCCT(15)-3'. Size selected >
1 kb for average insert length 1.7 kb. This is a primary
library, non-amplified. Library constructed by Life
Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
College of Medicine) and is available through Life
Technologies."
BASE COUNT 254 a 209 c 256 g 210 t
ORIGIN

```

```

Query Match 25.1%; Score 720; DB 13; Length 929;
Best Local Similarity 100.0%; Pred. No. 7.6e-26;
Matches 720; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2146 GGAATTCAGGGCAACAGGGAATTCAGGGCCATCATCGTGCAGAGAGAGAGAGGTGAA 2205
Db 1 GGAATTCAGGGCAACAGGGAATTCAGGGCCATCATCGTGCAGAGAGAGAGAGGTGAA 60
QY 2206 AAGGAGAACTCTGGTTCAGGAGTCCCATTTGGATTCAGAGGAGAAATCTGGGTGGATGCC 2265
Db 61 AAGGAGAACTCTGGTTCAGGAGTCCCATTTGGATTCAGAGGAGAAATCTGGGTGGATGCC 120
QY 2266 TTGATGGGGCCCGCAGGTCTCTAAGGGCAACCTGGGATCCAGGTCTCCAGGACCCCCA 2325
Db 121 TTGATGGGGCCCGCAGGTCTCTAAGGGCAACCTGGGATCCAGGTCTCCAGGACCCCCA 180
QY 2326 GTTTGGATGGGAAGCCCGGAAGAGATTTTCAGAACAAATTTATTCGACAAAGTTGGACA 2385
Db 181 GTTTGGATGGGAAGCCCGGAAGAGATTTTCAGAACAAATTTATTCGACAAAGTTGGACA 240
QY 2386 GATGTAATAGAGCCCGAGCTACAGTCTTACTTACAGAGTGGAGAAATAGAAATTTGGAT 2445
Db 241 GATGTAATAGAGCCCGAGCTACAGTCTTACTTACAGAGTGGAGAAATAGAAATTTGGAT 300
QY 2446 CATTCGCTGTCCCAACATGCTCCCGGGTATTCCTCGGCACTCGGTCCGATAGGCCCA 2505
Db 301 CATTCGCTGTCCCAACATGCTCCCGGGTATTCCTCGGCACTCGGTCCGATAGGCCCA 360
QY 2506 GAGGTCCTCCAGAGATTACCTGGTTTCCAGGAAGAGATGGTGTCTCTGGATTTAGTGGT 2565
Db 361 GAGGTCCTCCAGAGATTACCTGGTTTCCAGGAAGAGATGGTGTCTCTGGATTTAGTGGT 420
QY 2566 GTCCCTGACGTCCAGGTGTCCAGAGTAAAGGCTTACAGAGAGAAATGGGGAATAA 2625
Db 421 GTCCCTGACGTCCAGGTGTCCAGAGTAAAGGCTTACAGAGAGAAATGGGGAATAA 480
QY 2626 GGGAGCCAAAGGGTTGGGTTCTCTGGAGAAACAGGTCTCTCTGGTCCCGCAGGTCCAGAG 2685
Db 481 GGGAGCCAAAGGGTTGGGTTCTCTGGAGAAACAGGTCTCTCTGGTCCCGCAGGTCCAGAG 540
QY 2686 GGCCTCTCTGGATTAAGCAAAAGAGTCTCCAGGAGAGCCAGGTCTCTCTGSCAAAGAT 2745
Db 541 GGCCTCTCTGGATTAAGCAAAAGAGTCTCCAGGAGAGCCAGGTCTCTCTGSCAAAGAT 600
QY 2746 GGAGACCATGAAAACTCTGGAATCCAGGGCAACAGGCCCCCGAGGATCTCCGACCCA 2805
Db 601 GGAGACCATGAAAACTCTGGAATCCAGGGCAACAGGCCCCCGAGGATCTCCGACCCA 660
QY 2806 TCATATGTTTATGTTAATTTGGCAGAGAGATCCGTTTCAGAAAAGGACCAAACTATTAG 2865
Db 661 TCATATGTTTATGTTAATTTGGCAGAGAGATCCGTTTCAGAAAAGGACCAAACTATTAG 720

```



```

Db      1920  GGTAGTCCCGGTTTCAAGGAGAGCAGGATCCCTGTGTCTCCGGGCGAGGATGGAACA 1979
Qy      1792  CGGGAGAGCTGGATCCAGGATTTCTGGAACCGAGGATTAATGGGCCAAAGGGA 1851
Db      1980  CGGGAGAGCTGGATCCAGGATTTCTGGAACCGAGGATTAATGGGCCAAAGGGA 2039
Qy      1852  GAAATTTGGGCTCCAGGACAGCAAGAAAGGAGCCCAAGGATGCTGTTTAATG 1911
Db      2040  GAAATTTGGGCTCCAGGACAGCAAGAAAGGAGCCCAAGGATGCTGTTTAATG 2099
Qy      1912  GGAAGCAATGGCTCACAGGCCAGCTCTGGAACAACCGGATCTAAGGGAAGCAAGGTGAA 1971
Db      2100  GGAAGCAATGGCTCACAGGCCAGCTCTGGAACAACCGGATCTAAGGGAAGCAAGGTGAA 2159
Qy      1972  CTTGGAAATCAAGGATGCTGGGGCTTCTGGGCTCAAGGGAGAACCCAGGACACGGT 2031
Db      2160  CTTGGAAATCAAGGATGCTGGGGCTTCTGGGCTCAAGGGAGAACCCAGGACACGGT 2219
Qy      2032  TCCCGAGGAAACAGGATACATGGGTTTACCGGGATTCAGGAAAGGGGACAAA 2091
Db      2220  TCCCGAGGAAACAGGATACATGGGTTTACCGGGATTCAGGAAAGGGGACAAA 2279
Qy      2092  GGAATCAAGGTGAAAGGATTCAGGGTCAAGGTCAAAAGGAGAGAGTCAAGGAAAT 2151
Db      2280  GGAATCAAGGTGAAAGGATTCAGGGTCAAGGTCAAAAGGAGAGAGTCAAGGAAAT 2339
Qy      2152  CAGGGCAACAGGGAATTCAGGCCATCATGGTGCAAAAGGAGAGAGTCAAGGAAAT 2211
Db      2340  CAGGGCAACAGGGAATTCAGGCCATCATGGTGCAAAAGGAGAGAGTCAAGGAAAT 2399
Qy      2212  GAACCTGGTCTCGAGTGCCTATGGATCAAAAGGAGAAATCTGGGTGGATGGCTGATG 2271
Db      2400  GAACCTGGTCTCGAGTGCCTATGGATCAAAAGGAGAAATCTGGGTGGATGGCTGATG 2459
Qy      2272  GGGCCGCGAGTCTTAAGGGGCAACCTGGGGATCCAGGTCTCAGGAGCCCCAGGTTG 2331
Db      2460  GGGCCGCGAGTCTTAAGGGGCAACCTGGGGATCCAGGTCTCAGGAGCCCCAGGTTG 2519
Qy      2332  GATGGGAAGCCCGAGAGAGTTCAGAACAAATTTATTCGACAGTTTGACAGATGTA 2391
Db      2520  GATGGGAAGCCCGAGAGAGTTCAGAACAAATTTATTCGACAGTTTGACAGATGTA 2579
Qy      2392  ATAAGAGCCAGCTACAGTCTTACTTCAGAGTGGAGAAATTAAGAAATTTGATCATTC 2451
Db      2580  ATAAGAGCCAGCTACAGTCTTACTTCAGAGTGGAGAAATTAAGAAATTTGATCATTC 2639
Qy      2452  CTGTCCCAACATGGCTCCCGGGTATTCCTGGGCGACCTGGTCCGATAGCCAGAGGT 2511
Db      2640  CTGTCCCAACATGGCTCCCGGGTATTCCTGGGCGACCTGGTCCGATAGCCAGAGGT 2699
Qy      2512  CCCAGAGGATTAATCTGGTTTCCAGGAGAGATGGTGTCTCTGGATTAATGGTGTCT 2571
Db      2700  CCCAGAGGATTAATCTGGTTTCCAGGAGAGATGGTGTCTCTGGATTAATGGTGTCT 2759
Qy      2572  GGAGCTCCAGGTGTACAGGATTAAGAGGCTTACAGGAGAAATTCGGGAAAGGGAGC 2631
Db      2760  GGAGCTCCAGGTGTACAGGATTAAGAGGCTTACAGGAGAAATTCGGGAAAGGGAGC 2819
Qy      2632  CAAGGTTTGGGTATCTCTGAGAAACAAAGGTCTCTCTGGTCCCCAGGTCACAGGCCC 2691
Db      2820  CAAGGTTTGGGTATCTCTGAGAAACAAAGGTCTCTCTGGTCCCCAGGTCACAGGCCC 2879
Qy      2692  CTTGGATTAAGCAAGAGTCTCTCAGGAGACCCAGGTCTCTCTGGCAAGATGAGAC 2751
Db      2880  CTTGGATTAAGCAAGAGTCTCTCAGGAGACCCAGGTCTCTCTGGCAAGATGAGAC 2939
Qy      2752  CATGGAACCTCGAATCCAGGCGCAACAGGCCCCCAGGCTCTCTGGCAAGATGAGAC 2811
Db      2940  CATGGAACCTCGAATCCAGGCGCAACAGGCCCCCAGGCTCTCTGGCAAGATGAGAC 2999
Qy      2812  TGTTTTGTGTAATTCAGAGAGATCCGTTTCAGAAAGAGCAAACTATTAG 2865
Db      3000  TGTTTTGTGTAATTCAGAGAGATCCGTTTCAGAAAGAGCAAACTATTAG 3053

```

## RESULT 2

BX458795

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BX458795 Homo sapiens 1201 bp mRNA linear EST 22-MAY-2003  
 5-PRIME, mRNA sequence.  
 BX458795 Homo sapiens (human)  
 EST.  
 BX458795.1 GI:31033009  
 Homo sapiens (human)  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 1201)  
 Li,W.B., Gruber,C., Jessee,J. and Pclayes,D.  
 Full-length cDNA libraries and normalization  
 Unpublished  
 Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 EVRY cedex - France  
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
 Library was constructed by Life Technologies, a division of  
 Invitrogen. This sequence belongs to sequence cluster 3377.r For  
 more information about this cluster, see  
 http://www.genoscope.cns.fr/  
 cgi-bin/cluster.cgi?seq=CS0DE009AF04P1&cluster=3377.r. Contact :  
 Feng Liang Email : fliang@lifetech.com URL :  
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
 Faraday Avenue Genoscope sequence ID : CS0DE009AF04QP1.  
 Location/Qualifiers  
 1. 1201  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0DE009YK07"  
 /tissue\_type="PLACENTA"  
 /clone\_lib="Homo sapiens PLACENTA"  
 /notes="Vector: pCMVSPORT 6; 1st strand cDNA was primed  
 with a NotI-oligo(dT) primer. Five prime end enriched,  
 double-strand cDNA was digested with Not I and cloned into  
 the Not I and EcoRV sites of the pCMVSPORT 6 vector.  
 Library was not normalized."  
 350 a 223 c 387 g 212 t 29 others

## FEATURES

source

Query Match

Best Local Similarity

Matches

914; Conservative

1; Mismatches

30; Indels

11; Gaps

6;

Qy 1576

CcAGGATCAAGGGTGAATGGGTGCCAAAGGAGACAAAGGATCACTGGATTTATGGC

1635

Db 68

CcAGGATCAAGGGTGAATGGGTGCCAAAGGAGACAAAGGATCACTGGATTTATGGC

127

Qy 1636

AAAAAGGTCGAAAGGTGAAAAGGGGAATGCTGGCTTCCTGGCTCCCTGGACCTGCT

1695

Db 128

AAAAAGGTCGAAAGGTGAAAAGGGGAATGCTGGCTTCCTGGCTCCCTGGACCTGCT

187

Qy 1696

GGAGAACCAAGAGACATGGAAGGATGGAATTAATGGGTAGTCCCGTTTCAAGGAGAA

1755

Db 188

GGAGAACCAAGAGACATGGAAGGATGGAATTAATGGGTAGTCCCGTTTCAAGGAGAA

247

Qy 1756

GCAGGATCCCTGGTCTCCGGGCGAGGATGGACCGGGGAGAGCTGGATCCACGA

1815

Db 248

GCAGGATCCCTGGTCTCCGGGCGAGGATGGACCGGGGAGAGCTGGATCCACGA

307

Qy 1816

TTCTCTGGAAACCGAGGATTAATGGGCGAAAGGGGAATGGGGCTCCAGGACAGCAA

1875

Db 308

TTCTCTGGAAACCGAGGATTAATGGGCGAAAGGGGAATGGGGCTCCAGGACAGCAA

367

Qy 1876

GGAAAAAAGAGAGCCCAAGGATCCCTGGTCTTAAATGGGAAGCAATGGCTCACAGGCCAG

1925

Db 368

GGAAAAAAGAGAGCCCAAGGATCCCTGGTCTTAAATGGGAAGCAATGGCTCACAGGCCAG

427

Web site: <http://www.shgc.stanford.edu>  
 Contact: (Dickson, Mark) mcd@paxil.stanford.edu  
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAK plate: 32 Row: k Column: 15  
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 18780272  
 This clone has the following problem: retained intron.

## FEATURES

## source

1..4145  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4814199"  
 /tissue\_type="Brain, hippocampus"  
 /clone\_lib="NIH MGC 95"  
 /lab\_host="DH10B"  
 /note="Vector: pBluescript"

BASE COUNT 1275 a 787 c 1017 g 1066 t  
 ORIGIN

Query Match 98.3%; Score 2815.2; DB 11; Length 4145;  
 Best Local Similarity 99.3%; Pred. No. 0;  
 Matches 2854; Conservative 0; Mismatches 3; Indels 17; Gaps 2;

QY 1 ATGGCTCAGTATATACATTTCTCGCATGGTTTGGTCTGCTCTTCTCAGATTCGTG 60  
 DB 188 ATGGCTCAGTATATACATTTCTCGCATGGTTTGGTCTGCTCTTCTCAGATTCGTG 247

QY 61 TTAGCTCAAGATGGGAGTAAGATCAAGTTGCTGACTGCTCCGACAGATTAGTTTTC 120  
 DB 248 TTAGCTCAAGATGGGAGTAAGATCAAGTTGCTGACTGCTCCGACAGATTAGTTTTC 307

QY 121 ATCTTAGATGGCTCTTATAGTGTGGCCAGAAAACTTTGAAATAGTGAAGAAGTGGCT 180  
 DB 308 ATCTTAGATGGCTCTTATAGTGTGGCCAGAAAACTTTGAAATAGTGAAGAAGTGGCT 367

QY 181 GTCAATATCAAAAAAATTTGACATAGGCCCCAAGTTTATTCAGTTGGAGTGTCAA 240  
 DB 368 GTCAATATCAAAAAAATTTGACATAGGCCCCAAGTTTATTCAGTTGGAGTGTCAA 427

QY 241 TATAGTGACTACCTGTGCTGGAGATTCTCTCGGAAGCTATGATTGAGAGAACTTTG 300  
 DB 428 TATAGTGACTACCTGTGCTGGAGATTCTCTCGGAAGCTATGATTGAGAGAACTTTG 487

QY 301 ACGGTCAGTGGAAATCCATCTACTTATAGGAGGAAACACAAAGACAGGGAAGGCCATC 360  
 DB 488 ACGGTCAGTGGAAATCCATCTACTTATAGGAGGAAACACAAAGACAGGGAAGGCCATC 547

QY 361 CAGTTTGGCTCGATTACCTTTTGGCAAGTCTCAGATTCTGACTAAGATAGCAGTG 420  
 DB 548 CAGTTTGGCTCGATTACCTTTTGGCAAGTCTCAGATTCTGACTAAGATAGCAGTG 607

QY 421 GTACTTACGGATGGCAATCCCAAGATGAGCTCAAGATGCGACTCAAGCAGCAAGAT 480  
 DB 608 GTACTTACGGATGGCAATCCCAAGATGAGCTCAAGATGCGACTCAAGCAGCAAGAT 667

QY 481 AGTAAGATACATTATTGCTATTGTTGTTGTTTTCAGAACAGAGATCCGACATTAGA 540  
 DB 668 AGTAAGATACATTATTGCTATTGTTGTTGTTTTCAGAACAGAGATCCGACATTAGA 727

QY 541 GCTATTGCCAAACAGCCCTCGTCTACTTATGTTGTTTATGTTGAAGACTATATTGCAATA 600  
 DB 728 GCTATTGCCAAACAGCCCTCGTCTACTTATGTTGTTTATGTTGAAGACTATATTGCAATA 787

QY 601 TCCAAATAAGGGAAGTATGAGAGCAAGAACTTTGTGAAGAATCTGCTGTCCAAACAGCA 660  
 DB 788 TCCAAATAAGGGAAGTATGAGAGCAAGAACTTTGTGAAGAATCTGCTGTCCAAACAGCA 847

QY 661 ATTCCAGTGGCAGCTCGTGATGAAGGGGATTTGATATCTTTTAGGTTTAGATGTAAT 720

DB 848 ATCCAGTGGCAGCTCGTGATGAAGGGGATTTGATATTCTTTAGGTTTAGATGTAAT 907  
 QY 721 AAAAAGGTTAAGAAAAAGATACAGCTTTTACCAAAAAAGATATAAGGATATGAAGTAACA 780  
 DB 908 AAAAAGGTTAAGAAAAAGATACAGCTTTTACCAAAAAAGATATAAGGATATGAAGTAACA 967  
 QY 781 TCAAAGTTGATTATCAGAACTACACAGCAATGTTTTCCCAAGAGTCTTCTCCATCA 840  
 DB 968 TCAAAGTTGATTATCAGAACTACACAGCAATGTTTTCCCAAGAGTCTTCTCCATCA 1027  
 QY 841 TATGATTGTTGTTACTCTCAAGATTTAAAGTCAAGAAAAATTTGGGATTTATGAGAAATA 900  
 DB 1028 TATGATTGTTGTTACTCTCAAGATTTAAAGTCAAGAAAAATTTGGGATTTATGAGAAATA 1087  
 QY 901 TTAAGTATGATGAAGGCCACAAATAGCAGTTTACCTTAAATGGTGTGGCAAAAACTTTA 960  
 DB 1088 TTAAGTATGATGAAGGCCACAAATAGCAGTTTACCTTAAATGGTGTGGCAAAAACTTTA 1147  
 QY 961 TTATTTACAAACACAGCGTAATTAATGGCTCACAGTGGTACCTTTTGTAAACCTCTCAA 1020  
 DB 1148 TTATTTACAAACACAGCGTAATTAATGGCTCACAGTGGTACCTTTTGTAAACCTCTCAA 1207  
 QY 1021 GTTAAGAGCTTTGTTGATGAAGCTGGCACCAAAATTCGTCTCTTAGTAAACAGAAAGAT 1080  
 DB 1208 GTTAAGAGCTTTGTTGATGAAGCTGGCACCAAAATTCGTCTCTTAGTAAACAGAAAGAT 1267  
 QY 1081 GTGACTTTGTTATATTGATGACCAACAAATTTGAAACAAAGCCCTTACATCCAGTTTAGGG 1140  
 DB 1268 GTGACTTTGTTATATTGATGACCAACAAATTTGAAACAAAGCCCTTACATCCAGTTTAGGG 1327

QY 1141 ATCTTGATCATATGGCCAAAACCCAAATTTGAAATATTTCTGGAAAAAGAAACTGTTTCAG 1220  
 DB 1328 ATCTTGATCATATGGCCAAAACCCAAATTTGAAATATTTCTGGAAAAAGAAACTGTTTCAG 1387

QY 1201 TTTGATGTCCAAAAGTTGCGAATCTACTGTGACCCAGAAACAGAAACACCGGGAGACAGCA 1260  
 DB 1388 TTTGATGTCCAAAAGTTGCGAATCTACTGTGACCCAGAAACAGAAACACCGGGAGACAGCA 1447

QY 1261 TGTGAGATTCCTGGATTT-----TGCTTTAATGTCCTCCAGTGTAGGTTCACCT 1311  
 DB 1448 TGTGAGATTCCTGGATTTAATGAGAGTGGCTTAAATGGTCCAGTGTAGGTTCACCT 1307

QY 1311 CCAGCTCCCTGTAATTGCTCTCCGGAAAAACCGAGACTTCAAGGCCCCCAAGGTGACCT 1371  
 DB 1508 CCAGCTCCCTGTAATTGCTCTCCGGAAAAACCGAGACTTCAAGGCCCCCAAGGTGACCT 1367

QY 1372 GGACTGGCTGGAAACCTCTGGCTACCTGGACCAACCTGTCAGATGGTAAAGCTGGATAT 1431  
 DB 1568 GGACTGGCTGGAAACCTCTGGCTACCTGGACCAACCTGTCAGATGGTAAAGCTGGATAT 1619

QY 1432 CAGGGAATTGAGGAGACACAGGTTTCCAGGATCTCCAGGAATCAAGGAGTCCAGGA 1491  
 DB 1620 CAGGGAATTGAGGAGACACAGGTTTCCAGGATCTCCAGGAATCAAGGAGTCCAGGA 1679

QY 1492 CTACCAAGTTTACAAAGAGAAACAGGCGAGATGGTGACAAAGGTTGATCGTGGACTTCCT 1551  
 DB 1680 CTACCAAGTTTACAAAGAGAAACAGGCGAGATGGTGACAAAGGTTGATCGTGGACTTCCT 1739

QY 1552 GGTATTCTGGCTTCATGCGATGCCAGGATCAAGGGTCAAAATGGTCCCAAGGAGAC 1611  
 DB 1740 GGTATTCTGGCTTCATGCGATGCCAGGATCAAGGGTCAAAATGGTCCCAAGGAGAC 1799

QY 1612 AAAAGGATCACCTGGATTTTATGCAAAAAAGGTTGCAAAAGCTCAAAAGGGGAATGCTGGC 1671  
 DB 1800 AAAAGGATCACCTGGATTTTATGCAAAAAAGGTTGCAAAAGCTCAAAAGGGGAATGCTGGC 1859

QY 1672 TTCCCTGGCTCCCTGGACCTGTGGAGAACCCAGGAAGACATCGAAAGGATGSAATTAATG 1731  
 DB 1860 TTCCCTGGCTCCCTGGACCTGTGGAGAACCCAGGAAGACATCGAAAGGATGSAATTAATG 1919

QY 1732 GSTAGTCCCGGTTTCAAGGGAGAAAGCAGGATCCCTGCTGCTCCGGGGCAGGATGGAACA 1791

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2003, 19:09:03 ; Search time 5306 Seconds  
(without alignments)  
13123.318 Million cell updates/sec

Title: US-09-996-611b-5  
Perfect score: 2865  
Sequence: 1 atggctcactatattacatt.....gaaaggacaaactattag 2865

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562794

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*

1:	em_estba:*
2:	em_esthm:*
3:	em_estin:*
4:	em_estmu:*
5:	em_estov:*
6:	em_estpl:*
7:	em_estro:*
8:	em_htc:*
9:	gb_est1:*
10:	gb_est2:*
11:	gb_htc:*
12:	gb_est3:*
13:	gb_est4:*
14:	gb_est5:*
15:	em_estfun:*
16:	em_estom:*
17:	em_gss_hum:*
18:	em_gss_inv:*
19:	em_gss_pln:*
20:	em_gss_vrt:*
21:	em_gss_fun:*
22:	em_gss_mam:*
23:	em_gss_mus:*
24:	em_gss_pro:*
25:	em_gss_rod:*
26:	em_gss_phg:*
27:	em_gss_vri:*
28:	gb_gss1:*
29:	gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2815.2	98.3	4145	11 BC045597	BC045597 Homo sapi
2	825.6	28.8	1201	13 BX458795	BX458795 BX458795
3	720	25.1	929	13 BQ893743	BQ893743 AGENCOURT
4	436.8	15.2	909	13 BX431407	BX431407 BX431407

5	421.4	14.7	531	10 BE899929	BE899929 180456 MA
6	420.6	14.7	518	10 BE236040	BE236040 143702 MA
7	377	13.2	527	14 CA395709	CA395709 cs68a05.Y
8	355.2	12.4	670	13 BU250509	BU250509 60340456
9	344	12.0	744	10 BG699698	BG699698 602681559
10	340.2	11.9	760	13 BU108124	BU108124 60311461
11	309	10.8	327	14 N46880	N46880 YV58A01.r1
12	291.2	10.2	2668	11 AX028339	AX028339 Mus muscu
13	291.2	10.2	2996	11 AX011547	AX011547 Mus muscu
14	291.2	10.2	3035	11 BC038017	BC038017 Mus muscu
15	289.5	10.1	3405	11 AX031071	AX031071 Mus muscu
16	280.6	9.8	2826	11 AX030350	AX030350 Mus muscu
17	277.8	9.7	387	14 N85186	N85186 J2812F Huma
18	276	9.6	276	9 AI434106	AI434106 t131b11.x
19	262.8	9.2	494	10 AW62485	AW62485 EST374558
20	255.2	8.9	357	14 N86630	N86630 J9414F Huma
21	251	8.8	4784	11 AK041115	AK041115 Mus muscu
22	250.8	8.8	4782	11 AK048546	AK048546 Mus muscu
23	250.4	8.7	452	9 AA329319	AA329319 EST12913
24	249.4	8.7	4783	11 AK029212	AK029212 Mus muscu
25	248.8	8.7	5157	11 AK040971	AK040971 Mus muscu
26	247.2	8.6	4018	11 AK031163	AK031163 Mus muscu
27	240.2	8.4	944	13 BU903999	BU903999 AGENCOURT
28	240.2	8.4	5367	11 AK084803	AK084803 Mus muscu
29	238.4	8.3	481	9 AI127272	AI127272 q570c59.x
30	233	8.1	233	9 AA227783	AA227783 z157h04.r
31	231.8	8.1	464	14 T70851	T70851 yd15ell.r1
32	220.8	7.7	3512	11 AK080682	AK080682 Mus muscu
33	206.6	7.2	3113	11 AK028295	AK028295 Mus muscu
34	201	7.0	2931	11 AK019448	AK019448 Mus muscu
35	200.8	7.0	2513	11 BC030649	BC030649 Homo sapi
36	198	6.9	4288	11 BC039222	BC039222 Mus muscu
37	195	6.8	1552	11 AK038481	AK038481 Mus muscu
38	189.4	6.6	3649	11 BC035387	BC035387 Homo sapi
39	186.8	6.5	3699	11 AK075707	AK075707 Mus muscu
40	185.6	6.5	1443	11 BC008620	BC008620 Homo sapi
41	184.6	6.4	485	9 AA451616	AA451616 x43b04.r
42	184.6	6.4	3216	11 AK053733	AK053733 Mus muscu
43	178.2	6.2	4287	11 AK044870	AK044870 Mus muscu
44	171.6	6.0	1144	14 CD496495	CD496495 CDA22.F04
45	171.4	6.0	1190	11 AK012466	AK012466 Mus muscu

## ALIGNMENTS

RESULT 1	BC045597	4145 bp	mRNA	linear	HTC 04-MAR-2003				
LOCUS	BC045597	Homo sapiens, clone IMAGE:4814199, mRNA.							
DEFINITION	BC045597	Homo sapiens, clone IMAGE:4814199, mRNA.							
ACCESSION	BC045597.1	G1:28374134							
VERSION	BC045597.1	G1:28374134							
KEYWORDS	HTC								
SOURCE	Homo sapiens (human)								
ORGANISM	Homo sapiens								
REFERENCE	1	(bases 1 to 4145)							
AUTHORS	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.								
TITLE	Strausberg, R.								
JOURNAL	Direct Submission								
REMARK	Submitted (31-JAN-2003) National Institutes of Health, Mammalian Gene Collection (MGC); Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 1A03, Bethesda, MD 20892-2590, USA								
COMMENT	N-H-MGC Project URL: <a href="http://mgc.nci.nih.gov">http://mgc.nci.nih.gov</a> Contact: MGC help desk								

Qy	2813	GTTT	AGT	TAAT	TGCC	AGAG	AGAT	CGTT	CAGAAA	AGG	2852
Db	1375	GTTT <th>AGT</th> <th>TAAT</th> <th>TGCC</th> <th>AGAG</th> <th>AGAT</th> <th>CGTT</th> <th>CAGAAA</th> <th>AGG</th> <th>1414</th>	AGT	TAAT	TGCC	AGAG	AGAT	CGTT	CAGAAA	AGG	1414

Search completed: October 2, 2003, 19:29:40  
Job time : 724 secs

QY 541 GCTATTGCCAAAGCCTTCCTACTATGCTGTTTATGTTGGAGACTATATTGCAATA 600  
 DB |||||  
 QY 736 GCTATTGCCAAAGCCTTCCTACTATGCTGTTTATGTTGGAGACTATATTGCAATA 795  
 DB |||||  
 QY 601 TCCAAATTAAGGAAAGTGAAGAGCAAGAACTTTGTGGAAGATCTGTCTGTCACACGA 660  
 DB |||||  
 QY 796 TCCAAATTAAGGAAAGTGAAGAGCAAGAACTTTGTGGAAGATCTGTCTGTCACACGA 855  
 DB |||||  
 QY 661 ATTCCAGTGGCAGCTCGTGATGAAGGGGATTTGATATTTCTTTTAGGTTTAGATTAAT 720  
 DB |||||  
 QY 856 ATTCCAGTGGCAGCTCGTGATGAAGGGGATTTGATATTTCTTTTAGGTTTAGATTAAT 914  
 DB |||||  
 QY 721 AAAAAGTTTAAGAAAGATACAGCTTTACCAAAAAGATAAAGGATATGAAGTAACA 780  
 DB |||||  
 QY 915 AAAAGGTTTAAGAAAGATACAGCTTTACCAAAAAGATAAAGGATATGAAGTAACA 974  
 DB |||||  
 QY 781 TCAAAAGTTTATGATGAGAACTCACAAGCAATGTTTCCAGAAAGTCTTCTCCATCA 840  
 DB |||||  
 QY 975 TCAAAAGTTTATGATGAGAACTCACAAGCAATGTTTCCAGAAAGTCTTCTCCATCA 1034  
 DB |||||  
 QY 841 TATGTTATTTGCTGCTACTCAAGATTTAAAGTCAAGAAATTTGGGATTTATGGAGATA 900  
 DB |||||  
 QY 1035 TATGTTATTTGCTGCTACTCAAGATTTAAAGTCAAGAAATTTGGGATTTATGGAGATA 1094  
 DB |||||  
 QY 901 TTAAGTATTGATGAAGGCCA 921  
 DB |||||  
 QY 1095 TTAAGTATTGATGAAGGCCA 1115  
 DB |||||

## RESULT 15

ABK71703

ID ABK71703 standard; cDNA; 1414 BP.

XX AC ABK71703;

XX AC ABK71703;

XX 30-JUL-2002 (first entry)

XX Human dithp polynucleotide #169.

XX Human dithp polynucleotide #169.

KW Human; dithp; diagnostic and therapeutic polynucleotide; gene; ss; bone;  
 cell proliferative disorder; cancer; tumour; autoimmune disorder; brain;  
 inflammatory disorder; viral infection; bacterial infection; seizure;  
 fungal infection; parasitic infections; developmental disorder; breast;  
 endocrine disorder; metabolic disorder; neurological disorder; cervix;  
 gastrointestinal disorder; transport disorder; gene therapy; kidney;  
 adrenal gland; bone marrow; lung; ovary; pancreas; prostate; spleen;  
 skin; testis; thymus.

OS Homo sapiens.

XX WO200202754-A2.

XX 14-MAR-2002.

XX 29-AUG-2001; 2001WO-US27127.

XX 05-SEP-2000; 2000US-229747P.

XX 05-SEP-2000; 2000US-229748P.

XX 05-SEP-2000; 2000US-229749P.

XX 05-SEP-2000; 2000US-229750P.

XX 05-SEP-2000; 2000US-229751P.

XX 05-SEP-2000; 2000US-230583P.

XX 06-SEP-2000; 2000US-230584P.

XX 06-SEP-2000; 2000US-230514P.

XX 06-SEP-2000; 2000US-230515P.

XX 06-SEP-2000; 2000US-230517P.

XX 06-SEP-2000; 2000US-230518P.

XX 06-SEP-2000; 2000US-230519P.

XX 06-SEP-2000; 2000US-230595P.

XX 06-SEP-2000; 2000US-230597P.

XX 06-SEP-2000; 2000US-230598P.

XX 06-SEP-2000; 2000US-230599P.

XX 06-SEP-2000; 2000US-230599P.

PR 06-SEP-2000; 2000US-230610P.  
 PR 06-SEP-2000; 2000US-230865P.  
 PR 06-SEP-2000; 2000US-230988P.  
 PR 07-SEP-2000; 2000US-230991P.  
 PR 07-SEP-2000; 2000US-231163P.  
 PR 07-SEP-2000; 2000US-231167P.  
 XX (INCY-) INCYTE GENOMICS INC.  
 XX Stuart J, Lincoln SE, Altus CM, Dufour GE, Chalup MS, Hillman JL;  
 Jones AL, Yu CY, Wright RJ, Giezen D, Liu TP, Yap FE, Dahl CR;  
 Moniyam MG, Bradley DL, Rohatgi SD, Harris B, Roseberry AM;  
 Gerstin EH, Peralta CH, David MH, Panzer SR, Flores V, Daffo A;  
 Marwaha R, Chen AJ, Chang SC, Au AP, Inman RR;  
 WP: 2002-383054/41.  
 P-PSDB: AB360112.

An isolated polynucleotide useful in diagnostics and therapeutics -

Claim 1; Page 486; 686pp; English.

XX The invention relates to human diagnostic and therapeutic (dithp)  
 CC polynucleotides and their associated polypeptides (dithp polypeptides).  
 CC The sequences of the invention are used in the treatment and diagnosis of  
 CC cell proliferative disorders (e.g. atherosclerosis, cirrhosis), cancers  
 CC (e.g. tumours of the adrenal gland, bone, bone marrow, brain, breast,  
 CC cervix, kidney, lung, ovary, pancreas, prostate, skin, spleen, testis or  
 CC thymus), autoimmune/inflammatory disorders (e.g. asthma, bronchitis,  
 CC psoriasis, osteoporosis), viral infections, bacterial infections, fungal  
 CC infections, parasitic infections, developmental disorders (e.g. anaemia,  
 CC epilepsy), seizure disorders (e.g. thrombosis, aneurysm), metabolic disorders  
 CC (e.g. obesity, diabetes), neurological disorders (e.g. stroke,  
 CC amyotrophic lateral sclerosis, multiple sclerosis), gastrointestinal  
 CC disorders (e.g. ulcerative colitis, lysinuria) and transport disorders  
 CC (e.g. myotonic dystrophy, catatonias, peripheral neuropathy). Sequences  
 CC ABK71535-ABK71809 represent human dithp polynucleotides of the invention.  
 XX

SQ Sequence 1414 BP; 425 A; 267 C; 295 G; 427 T; 0 other;

Query Match 15.8%; Score 453.6; DB 24; Length 1414;  
 Best Local Similarity 99.1%; Pred No 2.7e-107;  
 Matches 456; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 2393 TAAGAGCCAGCTACCAAGTCTTACTTCAGAGTGAAGAAATTGATCATGGC 2452  
 DB |||||  
 QY 955 TGAAGCCAGCTACCAAGTCTTACTTCAGAGTGAAGAAATTGATCATGGC 1014  
 DB |||||  
 QY 2453 TGTCCCAACATGGCTCCCGGGTATTCTTGGGCCACCTGGTCCGATAGGCCAGAGGGTC 2512  
 DB |||||  
 QY 1015 TGTCCCAACATGGCTCCCGGGTATTCTTGGGCCACCTGGTCCGATAGGCCAGAGGGTC 1074  
 DB |||||  
 QY 2513 CCAGAGATTACCTGGTTTCCAGAGAGAGATGGTCTTCTGGATTAGTGGGTGTCCTG 2572  
 DB |||||  
 QY 1075 CCAGAGATTACCTGGTTTCCAGAGAGAGATGGTCTTCTGGATTAGTGGGTGTCCTG 1134  
 DB |||||  
 QY 2573 GACCTCAGGTGTCAGAGGATTAAAGGCTTACAGAGAGAAATGGGAAAAGGGAGCC 2632  
 DB |||||  
 QY 1135 GACCTCAGGTGTCAGAGGATTAAAGGCTTACAGAGAGAAATGGGAAAAGGGAGCC 1194  
 DB |||||  
 QY 2633 AAGGGTTTGGGTATCTCTGGAGAACAAAGTCTCTCTGGTCCCGCCAGGTCCAGAGGGCCCTC 2692  
 DB |||||  
 QY 1195 AAGGGTTTGGGTATCTCTGGAGAACAAAGTCTCTCTGGTCCCGCCAGGTCCAGAGGGCCCTC 1254  
 DB |||||  
 QY 2693 CTGGAATAGCAAGAGAGGTCTCTCCAGGAGACCCAGGTCTCCCTGGCAAGATGGAGACC 2752  
 DB |||||  
 QY 1255 CTGGAATAGCAAGAGAGGTCTCTCCAGGAGACCCAGGTCTCTCTGGCAAGATGGAGACC 1314  
 DB |||||  
 QY 2753 ATGGAAAACCTGGAAATCAAGAGGCAACCCAGGCCCCCAGGCATCTCCGACCCATCATAT 2812  
 DB |||||  
 QY 1315 ATGGAAAACCTGGAAATCAAGAGGCAACCCAGGCCCCCAGGCATCTCCGACCCATCATAT 1374  
 DB |||||

QY 961 TTATTACAAAC 974  
 DB 1156 TTATTACAAAC 1169

RESULT 14  
 AAS31132  
 ID AAS31132 standard; cDNA; 1115 BP.  
 XX  
 AC AAS31132;  
 XX  
 DT 04-DEC-2001 (first entry)  
 XX  
 DE Human diagnostic and therapeutic polynucleotide (DITHP) #147.  
 XX  
 KW Human; receptor; diagnostic; therapeutic; gene therapy; vaccine;  
 KW cell proliferative disorder; Crohn's disease; lymphoma; leukaemia;  
 KW acquired immune deficiency syndrome; AIDS; autoimmune disorder;  
 KW respiratory disorder; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200162927-A2.  
 XX  
 PD 30-AUG-2001.  
 XX  
 PF 21-FEB-2001; 2001WO-US06059.  
 XX  
 PR 24-FEB-2000; 2000US-0184693.  
 PR 24-FEB-2000; 2000US-0184697.  
 PR 24-FEB-2000; 2000US-0184698.  
 PR 24-FEB-2000; 2000US-0184768.  
 PR 24-FEB-2000; 2000US-0184769.  
 PR 24-FEB-2000; 2000US-0184770.  
 PR 24-FEB-2000; 2000US-0184771.  
 PR 24-FEB-2000; 2000US-0184772.  
 PR 24-FEB-2000; 2000US-0184773.  
 PR 24-FEB-2000; 2000US-0184774.  
 PR 24-FEB-2000; 2000US-0184776.  
 PR 24-FEB-2000; 2000US-0184777.  
 PR 24-FEB-2000; 2000US-0184797.  
 PR 24-FEB-2000; 2000US-0184813.  
 PR 24-FEB-2000; 2000US-0184837.  
 PR 24-FEB-2000; 2000US-0184841.  
 PR 24-FEB-2000; 2000US-0185213.  
 PR 24-FEB-2000; 2000US-0185216.  
 PR 12-MAY-2000; 2000US-0203785.  
 PR 15-MAY-2000; 2000US-0204226.  
 PR 16-MAY-2000; 2000US-0204525.  
 PR 16-MAY-2000; 2000US-0204821.  
 PR 16-MAY-2000; 2000US-0204908.  
 PR 16-MAY-2000; 2000US-0205232.  
 PR 17-MAY-2000; 2000US-0204815.  
 PR 17-MAY-2000; 2000US-0204863.  
 PR 17-MAY-2000; 2000US-0205221.  
 PR 17-MAY-2000; 2000US-0205285.  
 PR 17-MAY-2000; 2000US-0205286.  
 PR 17-MAY-2000; 2000US-0205287.  
 PR 17-MAY-2000; 2000US-0205323.  
 PR 17-MAY-2000; 2000US-0205324.  
 XX  
 PA (INCY-) INCYTE GENOMICS INC.  
 XX  
 PI Panzer SR, Spiro PA, Banville SC, Shah P, Chalup MS, Chang SC;  
 PI Chen A, D'Sa SA, Amesey S, Dahl CR, Dam TC, Daniels SE;  
 PI Dufour GE, Flores V, Fong WT, Greenawalt LB, Hillman JL, Jones AL;  
 PI Liu TF, Roseberry AM, Rosen BH, Russo FD, Stockdreher TK, Daffo A;  
 PI Wright RJ, Yap PE, Yu JY, Bradley DL, Bratcher SR, Chen W;  
 PI Cohen HJ, Hodgson DM, Lincoln SE, Jackson S;  
 XX  
 WIPI; 2001-502867/55.  
 DR P-PSDB; AAU19561.  
 XX

PT Polynucleotides encoding diagnostic and therapeutic proteins, e.g.  
 PT enzymes, hormones and receptors, useful in diagnostics and therapeutics  
 XX  
 PS Claim 1; Page 365; 522pp; English.  
 XX  
 CC The invention relates to polynucleotides (I) encoding diagnostic and  
 CC therapeutic (DITHP) polypeptides (II), which include e.g. enzymes,  
 CC and proteins involved in growth and development and receptors. (I) and  
 CC (II) may be used in the prevention, diagnosis and treatment of diseases  
 CC associated with inappropriate DITHP expression. For example, (I) and  
 CC (II) may be used to treat disorders associated with decreased polypeptide  
 CC expression by rectifying mutations or deletions in a patient's genome,  
 CC that affect the activity of the DITHPs, by expressing inactive proteins  
 CC or supplementing the patient's own production of them. (I) and (II)  
 CC may be used to treat diseases, for example, cell proliferative disorder,  
 CC Crohn's disease, acquired immune deficiency syndrome (AIDS), lymphoma,  
 CC leukaemia, autoimmune disorders, and respiratory disorders. Additionally,  
 CC (I) may be used to produce the DITHPs, by inserting the nucleic acids  
 CC into a host cell and culturing the cell to express the protein. (I) and  
 CC its complementary sequences may also be used as DNA probes in diagnostic  
 CC assays to detect and quantitate the presence of similar nucleic acids in  
 CC samples, and therefore which patients may be in need of restorative  
 CC therapy. (II) may also be used as antigens in the production of  
 CC antibodies against DITHPs and in assays to identify modulators of DITHP  
 CC expression and activity. The anti-DITHP antibodies and antagonists may  
 CC also be used to down regulate expression and activity. The anti-DITHP  
 CC antibodies may also be used as diagnostic agents for detecting the  
 CC presence of DITHPs in samples (e.g. by enzyme linked immunosorbant  
 CC assay (ELISA)). AAS30986-AAS31196 represent human diagnostic and  
 CC therapeutic (DITHP) polynucleotides of the invention.  
 XX  
 SQ Sequence 1115 BP; 344 A; 215 C; 259 G; 296 T; 1 other;

Query Match 31.7%; Score 908; DB 22; Length 1115;  
 Best Local Similarity 99.8%; Pred. No. 1.6e-225;  
 Matches 919; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 ATGGCTCACTATATTACATTTCTCTGATGTTTGGTGGTCTTCTTCTCAGAAATTCGTG 60  
 DB 196 ATGGCTCACTATATTACATTTCTCTGATGTTTGGTGGTCTTCTTCTCAGAAATTCGTG 255  
 QY 61 TTAGCTGAAGATGGGAAGTAAGATCAAGTTGTGTAAGTCTCCGACAGATTTAGTTTTC 120  
 DB 256 TTAGCTGAAGATGGGAAGTAAGATCAAGTTGTGTAAGTCTCCGACAGATTTAGTTTTC 315  
 QY 121 ATCTTAGATGCTCTTATAGTGTGGCCAGAAACTTTGAATAGTGAAGAGTGGCTT 180  
 DB 316 ATCTTAGATGCTCTTATAGTGTGGCCAGAAACTTTGAATAGTGAAGAGTGGCTT 375  
 QY 181 GTCAATATCAAAAAAATTTGACATAGGCGCGAAGTTTATTCAGTTGGAGTGTTCAA 240  
 DB 376 GTCAATATCAAAAAAATTTGACATAGGCGCGAAGTTTATTCAGTTGGAGTGTTCAA 435  
 QY 241 TATAGTGACTACCTGTGCTGGAGATTCCTCTCGAAGCTATGATTCAGAGACAAATTG 300  
 DB 436 TATAGTGACTACCTGTGCTGGAGATTCCTCTCGAAGCTATGATTCAGAGACAAATTG 495  
 QY 301 ACGGACGAGTGGAAATCCATCTACTTATAGGAGGAAACACAAAGACAGGAGGCCATC 360  
 DB 496 ACGGACGAGTGGAAATCCATCTACTTATAGGAGGAAACACAAAGACAGGAGGCCATC 555  
 QY 361 CAGTTTGGCTCGAATACCTTTTTCGCAAGTCTCTACAGTATTCAGTAAAGATAGCAGTG 420  
 DB 556 CAGTTTGGCTCGAATACCTTTTTCGCAAGTCTCTACAGTATTCAGTAAAGATAGCAGTG 615  
 QY 421 GTACTTACGATGGGAAATCCCAAGATGACGTCAGGATGCGAGCTCAACGACAGAGAT 480  
 DB 616 GTACTTACGATGGGAAATCCCAAGATGACGTCAGGATGCGAGCTCAACGACAGAGAT 675  
 QY 481 AGTAAGATACATTTATTTGCTATTTGGTGTTCAGAAACAGAGATCCCAACTTACA 540  
 DB 676 AGTAAGATACATTTATTTGCTATTTGGTGTTCAGAAACAGAGATCCCAACTTACA 735

```
OS Homo sapiens.
XX WO200240715-A2.
XX 23-MAY-2002.
XX
XX 06-SEP-2001; 2001WO-US27628.
XX
XX 06-SEP-2000; 2000US-230505P.
XX 06-SEP-2000; 2000US-230514P.
XX 06-SEP-2000; 2000US-230515P.
XX 06-SEP-2000; 2000US-230517P.
XX 06-SEP-2000; 2000US-230518P.
XX 06-SEP-2000; 2000US-230519P.
XX 06-SEP-2000; 2000US-230595P.
XX 06-SEP-2000; 2000US-230597P.
XX 06-SEP-2000; 2000US-230598P.
XX 06-SEP-2000; 2000US-230599P.
XX 06-SEP-2000; 2000US-230610P.
XX 06-SEP-2000; 2000US-230655P.
XX 06-SEP-2000; 2000US-230685P.
XX 06-SEP-2000; 2000US-230989P.
XX 07-SEP-2000; 2000US-230951P.
XX 07-SEP-2000; 2000US-231163P.
XX 07-SEP-2000; 2000US-231167P.
XX
XX (INCY-) INCYTE GENOMICS INC.
XX
XX Jackson S, Lincoln SE, Altus CM, Dufour GE, Chalup MS, Hillman JL,
XX Jones AL, Yu JY, Wright RJ, Gietzen D, Liu TF, Yap PE, Dahl CR,
XX Momiyama MG, Bradley DL, Rohatgi SD, Harris B, Roseberry AM,
XX Gerstin EH, Peralta CH, David MH, Panzer SR, Flores V, Daffo A,
XX Marwaha R, Chen AJ, Chang SC, Au AP, Inman RR;
XX
XX WPI; 2002-527544/56.
XX P-PSDB; ABP51293.
XX
XX Novel human disease detection and treatment polypeptide, useful in
XX diagnosis, prevention or treatment of cell proliferative disorders e.g.
XX arteriosclerosis, cirrhosis and an autoimmune/inflammatory disorder
XX e.g. AIDS
XX
XX Claim 1; Page 341; 618pp; English.
XX
XX The invention relates to an isolated human disease detection and
XX treatment (MDPT) polypeptide (I) selected from a polypeptide having a
XX sequence selected from 254 sequences (ABP51231-ABP51484) given in the
XX specification, a naturally occurring polypeptide comprising a sequence
XX having at least 90% identity to (I) or a biologically active or
XX immunogenic fragment of (I). (I) is useful for screening a compound for
XX effectiveness as an agonist or antagonist, for screening a compound that
XX specifically binds (I) or modulates the activity of (I), and for
XX preparing a polyclonal or monoclonal antibody by hybridoma technology.
XX Nucleic acids (II) (ABQ72449-ABQ72700) encoding (I) are useful for
XX screening a compound for effectiveness in altering expression of a target
XX polynucleotide comprising. Oligonucleotides and antibodies are useful for
XX detecting MDDT in a sample or for assessing toxicity of a test compound,
XX in a diagnostic test for a condition or a disease associated with the
XX expression of MDDT in a biological sample, for detecting (I) in a sample,
XX and for purifying (I) from a sample. A composition comprising (I), an
XX agonist or antagonist is useful for treating a disease or condition
XX associated with decreased or increased expression of functional MDDT.
XX (I) or (II) are useful for diagnosing, treating or preventing disorders
XX selected with aberrant expression of MDDT, where the disorders are
XX arteriosclerosis, cirrhosis, psoriasis, and cancer and an
XX autoimmune/inflammatory disorder such as AIDS, Addison's disease,
XX allergy, anaemia, asthma, atherosclerosis, gout, multiple sclerosis or
XX rheumatoid arthritis. (II) are useful for creating knockin humanised
XX animals or transgenic animals to model human diseases, in somatic or
XX germ-line gene therapy, to generate a transcript image of a tissue or cell
XX type, for detecting differences in the chromosomal location due to
XX translocation or inversion among normal, carrier or affected individuals
```

```
CC and as hybridisation probes for mapping naturally occurring genomic
CC sequences.
XX
XX Sequence 1169 BP: 364 A; 224 C; 267 G; 314 T; 0 Other;
XX
XX Query Match: 33.9%; Score 972.4; DB 24; Length 1169;
XX Best Local Similarity 99.9%; P-val. No. 3e-242;
XX Matches 973; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
Qy 1 ATGGCTCACTATATTACATTTCTCTGCATGGTTTGGTGGTCTCTCTTCAGAAATCTGTG 60
Db 196 ATGGCTCACTATATTACATTTCTCTGCATGGTTTGGTGGTCTCTCTTCAGAAATCTGTG 255
Qy 61 TTAGCTGAGATGGGGAAGTAAGATCAAGTTGTCTGCTACTGCTCCGACAGATTTAGTTTC 120
Db 256 TTAGCTGAGATGGGGAAGTAAGATCAAGTTGTCTGCTACTGCTCCGACAGATTTAGTTTC 315
Qy 121 ATCTTAGATGGCTCTTATAGTGTGGCCGAGAAACTTTGAAATAGTGAAGTGGCTT 180
Db 316 ATCTTAGATGGCTCTTATAGTGTGGCCGAGAAACTTTGAAATAGTGAAGTGGCTT 375
Qy 181 GTCAATATCAAAAAAATTTGACATAGGGCCGAGTTTATTCNAGTTGGAGTGTCAA 240
Db 376 GTCAATATCAAAAAAATTTGACATAGGGCCGAGTTTATTCNAGTTGGAGTGTCAA 435
Qy 241 TATAGTGAATACCTCTGTCTGGAGATTCCTCTGGAAAGCTATGATTCAGGAGAACATTTG 300
Db 436 TATAGTGAATACCTCTGTCTGGAGATTCCTCTGGAAAGCTATGATTCAGGAGAACATTTG 495
Qy 301 ACGGACAGTGAATCCATCTCTACTTAGGAGGAAACACAAAGACAGGAGGCATC 360
Db 496 ACGGACAGTGAATCCATCTCTACTTAGGAGGAAACACAAAGACAGGAGGCATC 555
Qy 361 CAGTTTGGCTCGATTAACCTTTTGGCAAGTCTCTCAGATTTCTGACTAAGATAGCAGTG 420
Db 556 CAGTTTGGCTCGATTAACCTTTTGGCAAGTCTCTCAGATTTCTGACTAAGATAGCAGTG 515
Qy 421 GTACTTACGGATGGCAATCCCAAGATGACGTCAAGGATGCGAGCTCAACAGCAGAGAT 480
Db 616 GTACTTACGGATGGCAATCCCAAGATGACGTCAAGGATGCGAGCTCAACAGCAGAGAT 675
Qy 481 AGTAAGATACATTAATTCCTATTTGGTGTGGTTCAGAAACAGAGATGCCGAATTAGA 540
Db 676 AGTAAGATACATTAATTTGGTGTGGTTCAGAAACAGAGATGCCGAATTAGA 735
Qy 541 GCTATTGCCAAAGGCTTCTGCTACTTATGTGTTTATGTGSAAGACTATATTGCAATA 600
Db 736 GCTATTGCCAAAGGCTTCTGCTACTTATGTGTTTATGTGSAAGACTATATTGCAATA 795
Qy 601 TCCAAATTAAGGGAAGTGATGAAGCAGAAACTTTGTGAAGAATCTGTCTGTCCACACGA 660
Db 796 TCCAAATTAAGGGAAGTGATGAAGCAGAAACTTTGTGAAGAATCTGTCTGTCCACACGA 855
Qy 661 ATTCCAGTGGCAGCTCGTGATGAAGGGGATTTGATTAATCTTTAGTGTAGATGTAAT 720
Db 856 ATTCCAGTGGCAGCTCGTGATGAAGGGGATTTGATTAATCTTTAGTGTAGATGTAAT 915
Qy 721 AAAAAGGTTAAGAAAAAATACAGCTTTTCCACAAAAAAGATATAAGATATGAAGTAACA 780
Db 916 AAAAAGGTTAAGAAAAAATACAGCTTTTCCACAAAAAAGATATAAGATATGAAGTAACA 975
Qy 781 TCAAAAGTTGATTTATCAGAACTCAGAACTATGTTTCCGAGAGGCTTCTCTCATCA 840
Db 976 TCAAAAGTTGATTTATCAGAACTCAGAACTATGTTTCCGAGAGGCTTCTCTCATCA 1035
Qy 841 TATGTTTGTGTCTTACTCAAGATTTAAAGTCAAGAAATTTGGATTTATGAGAGAA 900
Db 1036 TATGTTTGTGTCTTACTCAAGATTTAAAGTCAAGAAATTTGGATTTATGAGAGAA 1095
Qy 901 TTAACCTTATGAGGAGGCGCACAAATAGCAGTTTACCTTAAATGGTGTGCACAAAACTT 960
Db 1096 TTAACCTTATGAGGAGGCGCACAAATAGCAGTTTACCTTAAATGGTGTGCACAAAACTT 1155
```

PR 07-SEP-2000; 2000US-231163P.  
 PR 07-SEP-2000; 2000US-231167P.  
 XX (INCY-) INCYTE GENOMICS INC.  
 PA Jackson S, Linceln SE, Altus CM, Dufour GE, Chalup MS, Hillman JL;  
 PI Jones AL, Yu JY, Wright RJ, Gietzen D, Liu TF, Yap PE, Dahl CR;  
 PI Moniyama MG, Bradley DL, Rohatgi SD, Harris B, Roseberry AM;  
 PI Gerstin EH, Peralta CH, David MH, Panzer SR, Flores V, Daffo A;  
 PI Marwaha R, Chen AJ, Chang SC, Au AP, Inman RR;  
 XX WPI: 2002-527544/56.  
 DR P-PSDB; ABP51423.  
 XX  
 PT Novel human disease detection and treatment polypeptide, useful in  
 PT diagnosis, prevention or treatment of cell proliferative disorders e.g.  
 PT arteriosclerosis, cirrhosis and an autoimmune/inflammatory disorder  
 PT e.g. AIDS  
 PT  
 PS Claim 1; Page 409-410; 618pp; English.  
 XX  
 CC The invention relates to an isolated human disease detection and  
 CC treatment (MDPT) polypeptide (I) selected from a polypeptide having a  
 CC sequence selected from 254 sequences (ABP51231-ABP51484) given in the  
 CC specification, a naturally occurring polypeptide comprising a sequence  
 CC having at least 90% identity to (I) or a biologically active or  
 CC immunogenic fragment of (I). (I) is useful for screening a compound for  
 CC effectiveness as an agonist or antagonist, for screening a compound that  
 CC specifically binds (I) or modulates the activity of (I), and for  
 CC preparing a polyclonal or monoclonal antibody by hybridoma technology.  
 CC Nucleic acids (II) (ABQ72449-ABQ72700) encoding (I) are useful for  
 CC screening a compound for effectiveness in altering expression of a target  
 CC polynucleotide comprising. Oligonucleotides and antibodies are useful for  
 CC detecting MDPT in a sample or for assessing toxicity of a test compound,  
 CC in a diagnostic test for a condition or a disease associated with the  
 CC expression of MDPT in a biological sample, for detecting (I) in a sample,  
 CC and for purifying (I) from a sample. A composition comprising (I), an  
 CC agonist or antagonist is useful for treating a disease or condition  
 CC associated with decreased or increased expression of functional MDPT.  
 CC (I) or (II) are useful for diagnosing, treating or preventing disorders  
 CC associated with aberrant expression of MDPT, where the disorders are  
 CC selected from a cell proliferative disorder such as arteriosclerosis,  
 CC cirrhosis, hepatitis, psoriasis, and cancer and an  
 CC autoimmune/inflammatory disorder such as AIDS, Addison's disease,  
 CC allergy, anaemia, asthma, atherosclerosis, gout, multiple sclerosis or  
 CC rheumatoid arthritis. (II) are useful for creating knockin humanised  
 CC animals or transgenic animals to model human diseases, in somatic or  
 CC germline gene therapy, to generate a transcript image of a tissue or cell  
 CC type, for detecting differences in the chromosomal location due to  
 CC translocation or inversion among normal, carrier or affected individuals  
 CC and as hybridisation probes for mapping naturally occurring genomic  
 CC sequences.  
 XX  
 SQ Sequence 1184 BP; 368 A; 227 C; 272 G; 317 T; 0 other;  
 Query Match 34.0%; Score 974; DB 24; Length 1184;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-242;  
 Matches 974; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ATGGCTCACTATATACATTTCTCTGCGATGGTTTTGGTGGTCTTCTTCCAGAAATTCGTG 60  
 DB 211 ATGGCTCACTATATACATTTCTCTGCGATGGTTTTGGTGGTCTTCTTCCAGAAATTCGTG 270  
 QY 61 TTAGCTGAAGATGGGAAGTAAGATCAAGTTGCTGCTACTGCTCGACAGATTAGTTTTC 120  
 DB 271 TTAGCTGAAGATGGGAAGTAAGATCAAGTTGCTGCTACTGCTCGACAGATTAGTTTTC 330  
 QY 121 ATCTTAGATGGCTCTTATAGTTTGGCCACAGAAACTTGAATAGTGAAGAAAGTGGCTT 180  
 DB 331 ATCTTAGATGGCTCTTATAGTTTGGCCACAGAAACTTGAATAGTGAAGAAAGTGGCTT 390  
 QY 181 GTCAATATCAAAAAAATTTGACATAGGGCCGAGGTTTATTCAGTTGAGTGGTTCAA 240

DB 391 GTCAATATCAAAAAAATTTGACATAGGGCCGAGGTTTATTCAGTTGAGTGGTTCAA 450  
 QY 241 TATAGTGACTACCTCTGCTGGAGATTCTCTGGAGATGATGATTCAGGAGAAATTTG 300  
 DB 451 TATAGTGACTACCTCTGCTGGAGATTCTCTGGAGATGATGATTCAGGAGAAATTTG 510  
 QY 301 ACGGACGACGTGAAATCCCATACCTCTACTTAGGAGGAAACACAAAGACAGGAGGCCATC 360  
 DB 511 ACGGACGACGTGAAATCCCATACCTCTACTTAGGAGGAAACACAAAGACAGGAGGCCATC 570  
 QY 361 CAGTTTCGGCTCGATTACCTTTTGGCAAGTCTCTACGATTTCTGACTAAGATACGAGTG 420  
 DB 571 CAGTTTCGGCTCGATTACCTTTTGGCAAGTCTCTACGATTTCTGACTAAGATACGAGTG 630  
 QY 421 GTACTTACGATGCGAAATCCCAAGATGACGTCAGGATGACCTCAACGACAGAGAT 480  
 DB 631 GTACTTACGATGCGAAATCCCAAGATGACGTCAGGATGACCTCAACGACAGAGAT 690  
 QY 481 AGTAAGATAACATTTATTTGCTATTGCTTGGTTTGGTTTCAAAAACAGAGATGCCGAACTT 540  
 DB 591 AGTAAGATAACATTTATTTGCTATTGCTTGGTTTGGTTTCAAAAACAGAGATGCCGAACTT 750  
 QY 541 GCATTCCCAACAGGCTTCTGCTACTTATGTTTATGTTGAAAGACTATATTGCAATA 600  
 DB 751 GCATTCCCAACAGGCTTCTGCTACTTATGTTTATGTTGAAAGACTATATTGCAATA 810  
 QY 601 TCCAAATTAAGGGAAGTGAAGACAGAACTTTGTGAAGAACTCTGCTGTCTCCACACCA 660  
 DB 811 TCCAAATTAAGGGAAGTGAAGACAGAACTTTGTGAAGAACTCTGCTGTCTCCACACCA 870  
 QY 661 ATTCCAGTGGCAGCTGTGATGAAAGGGGATTTGATPATCTTTTAGTTTAGATGTAAAT 720  
 DB 871 ATTCCAGTGGCAGCTGTGATGAAAGGGGATTTGATPATCTTTTAGTTTAGATGTAAAT 930  
 QY 721 AAAAGGTTAAGAAAGATACAGCTTCCAGAAAGATTAAGAGATTAAGATTAAGTAA 780  
 DB 931 AAAAGGTTAAGAAAGATACAGCTTCCAGAAAGATTAAGAGATTAAGATTAAGTAA 990  
 QY 781 TCAAAAGTTGATTTATCAGAACTCAAGCAAGCAATGTTTCCAGAAAGTCTTCTCCATCA 840  
 DB 991 TCAAAAGTTGATTTATCAGAACTCAAGCAAGCAATGTTTCCAGAAAGTCTTCTCCATCA 1050  
 QY 841 TATGATTTGCTCTACTCAAGATTTAAAGTCAAGAAATTTGGGATTTAGGAGAAATA 900  
 DB 1051 TATGATTTGCTCTACTCAAGATTTAAAGTCAAGAAATTTGGGATTTAGGAGAAATA 1110  
 QY 901 TTAAGTATTTGATGAGGAGGACAAATAGCAGTTTACCTTAAATGCTGTGACAAATCTT 960  
 DB 1111 TTAAGTATTTGATGAGGAGGACAAATAGCAGTTTACCTTAAATGCTGTGACAAATCTT 1170  
 QY 961 TTATTTTCAACAAAC 974  
 DB 1171 TTATTTTCAACAAAC 1184  
 RESULT 13  
 ABQ72511  
 ID ABQ72511 standard; cDNA; 1169 BP.  
 XX  
 AC ABQ72511;  
 XX  
 DT 03-SEP-2002 (first entry)  
 XX  
 DE Human MDDT encoding cDNA SEQ ID NO 63.  
 XX  
 KW Human; MDDT; disease detection and treatment molecule polynucleotide;  
 KW proliferative disorder; hepatitis; psoriasis; cancer; AIDS;  
 KW autoimmune disorder; inflammatory disorder; allergy; multiple sclerosis;  
 KW rheumatoid arthritis; transgenic; gene therapy; antiarteriosclerotic;  
 KW hepatotropic; antiinflammatory; antipsoriatic; cytostatic; anti-HIV;  
 KW antiallergic; antianemic; antiasthmatic; antiatherosclerotic; anti-gout;  
 KW neuroprotective; antirheumatic; antiarthritic; gene; ss.  
 XX



PT Novel nucleic acids and polypeptides, useful for treating disorders  
PT such as central nervous system injuries -  
XX  
XX  
XX Claim 1; SEQ ID NO 4177; 10078pp; English.  
XX  
CC The invention relates to human nucleic acids (AA157798-AA161369) and  
CC the encoded polypeptides (AA138642-AA142213) with nootropic,  
CC immunosuppressant and cytotostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: Immune system suppression,  
CC activation/inhibition activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemias and  
CC C.N.S disorders.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification.  
XX  
XX Sequence 2209 BP; 674 A; 413 C; 544 G; 577 T; 1 other;  
SQ  
Query Match 36.1%; Score 1039.8; DB 22; Length 2209;  
Best Local Similarity 98.1%; Pred. No. 1.3e-259;  
Matches 1107; Conservative 0; Mismatches 12; Indels 10; Gaps 5;  
QY 1747 AAGGAGACGAGATCCCTGCTGCTCCGGGACGAGTGAACAGGGGAGAG-----C 1801  
DB 4 AAGGAGAGACGAGATCCCTGCTGCTCCGGGACGAGTGAACAGGGGAGAGAACCC 63  
QY 1802 CTGGAATCCCGAGATT--CCTGGAACCGAGATTAAATGGG-CCAAAGAGGAGAAATTG 1858  
DB 64 CTGGAATCCCGAGATTTCCTGTGAACCGAGATTAAATGGGCCCAAGAGGAGAAATTG 123  
QY 1859 GGCCT-CCAGGACAGCAAGGAAAAAGAGAGCCCGAGGATG-CCTGGTTTAAATGGGAAG 1916  
DB 124 GGCCTCCAGGACAGCAAGGAAAAAGAGAGCCCGAGGATGCTGTTTAAATGGGAAG 183  
QY 1917 CAATGGCTACACAGGCGACCTGGAAACCGGATCTAAGGGAAGCAAGGTCAACCTGG 1976  
DB 184 CAATGGCTACACAGGCGACCTGGAAACCGGATCTAAGGGAAGCAAGGTCAACCTGG 243  
QY 1977 AATCAAGGATGCTGGGGCTTCTGGGTCAAGGAGAAACAGGAGCAACGGTTCCCG 2036  
DB 244 AATCAAGGATGCTGGGGCTTCTGGGTCAAGGAGAAACAGGAGCAACGGTTCCCG 303  
QY 2037 AGGAGAACGAGATACATGGTTTACCGGATTCAGGAAAGGAGGAGCAAGGAA 2096  
DB 304 AGGAGAACGAGATACATGGTTTACCGGATTCAGGAAAGGAGGAGCAAGGAA 363  
QY 2097 TCAAGGTGAAAAAGGTATTCAAGGTCAAAAGGAGAGAAAAATGAAGACAGGAAATCCAGG 2156  
DB 364 TCAAGGTGAAAAAGGTATTCAAGGTCAAAAGGAGAGAAAAATGAAGACAGGAAATCCAGG 423  
QY 2157 GCACAGGGAATTCAGAGCATCATGTGCAAAAGAGAGAGAGGTGAAGAGGAGAAC 2216  
DB 424 GCACAGGGAATTCAGAGCATCATGTGCAAAAGAGAGAGAGGTGAAGAGGAGAAC 483  
QY 2217 TGGTGTCCGAGGTGCCATTGGATCAAAAGAGAGATCTGGGTGGATGGCTTCATCGGGCC 2276  
DB 484 TGGTGTCCGAGGTGCCATTGGATCAAAAGAGAGATCTGGGTGGATGGCTTCATCGGGCC 543  
QY 2277 CGCAGGTCTTAAGGGCAACCTGGGGATCCAGTCTCAGGGAGCCCGAGGTTTGGATGG 2336  
DB 544 CGCAGGTCTTAAGGGCAACCTGGGGATCCAGTCTCAGGGAGCCCGAGGTTTGGATGG 603  
QY 2337 GAAGCCCGAGAGAGTTTTCAGAACAAATTTATTCACAGATTTCACAGATGTAATAAG 2366  
DB 604 GAAGCCCGAGAGAGTTTTCAGAACAAATTTATTCACAGATTTCACAGATGTAATAAG 663  
QY 2397 AGCCACGCTACCAGTCTTACTTCAGAGTGAAGAAATTAGAAATTTGATCATCTGCTGTC 2456

DB 564 AGCCAGCTACAGTCTTACTTCAGAGTGAAGAAATTAGAAATTTGATCATCTGCTGTC 723  
QY 2457 CCAACATGCTCCCGGATATTCCTGGGACCTGGTCCGATAGGCCAGAGGTTCCAG 2516  
DB 724 CCAACATGCTCCCGGATATTCCTGGGACCTGGTCCGATAGGCCAGAGGTTCCAG 783  
QY 2517 AGGATTACCTGTTTCCCAAGAGAGATGGTCTTCTGATTTAGTGGGTGCTCCCTGGACG 2576  
DB 784 AGGATTACCTGTTTCCCAAGAGAGATGGTCTTCTGATTTAGTGGGTGCTCCCTGGACG 843  
QY 2577 TCCAGTGTCCAGAGATTAAAAAGGCTACCAAGGAAGAAATGGGAAAAAGGAGCCAAAG 2636  
DB 844 TCCAGTGTCCAGAGATTAAAAAGGCTACCAAGGAAGAAATGGGAAAAAGGAGCCAAAG 903  
QY 2637 GTTGGGTATCTCGAGACACAGGTCTCTGCTGGTCCGACAGTCCAGAGGCGCTCTCTGG 2696  
DB 904 GTTGGGTATCTCGAGAACAAAGGTCTCTGCTGGTCCGACAGTCCAGAGGCGCTCTCTGG 963  
QY 2697 AATAAGCAAGAGGTCTCTCCAGGAGACCCAGTCTCTGCTGGTCCGACAGTCCAGAGGCGCTCTCTGG 2755  
DB 964 AATAAGCAAGAGGTCTCTCCAGGAGACCCAGTCTCTGCTGGTCCGACAGTCCAGAGGCGCTCTCTGG 1023  
QY 2757 AATAAGCAAGAGGTCTCTCCAGGAGACCCAGTCTCTGCTGGTCCGACAGTCCAGAGGCGCTCTCTGG 2816  
DB 1024 AATAAGCAAGAGGTCTCTCCAGGAGACCCAGTCTCTGCTGGTCCGACAGTCCAGAGGCGCTCTCTGG 1083  
QY 2817 TACTGTAAATTCGCAAGAGAGATCCGTTTCAGAAAGGAGCCAAACTATTAG 2865  
DB 1084 TACTGTAAATTCGCAAGAGAGATCCGTTTCAGAAAGGAGCCAAACTATTAG 1132  
RESULT 12  
ABQ72640  
ID ABQ72640 standard; cDNA: 1184 BP.  
XX AC ABQ72640;  
XX DT 03-SEP-2002 (first entry)  
XX DE Human MDDT encoding cDNA SEQ ID NO 192.  
XX KW Human: MDDT; disease detection and treatment molecule polynucleotide;  
KW proliferative disorder; hepatitis; psoriasis; cancer; AIDS;  
KW autoimmune disorder; inflammatory disorder; allergy; multiple sclerosis;  
KW rheumatoid arthritis; transgenic; gene therapy; antiatherosclerotic;  
KW hepatotopic; antiinflammatory; antipsoriatic; cytostatic; anti-HIV;  
KW antiallergic; antianemic; antiasthmatic; antithrombotic; antidiabetic;  
KW neuroprotective; antirheumatic; antiarthritic; gene; ss.  
XX OS Homo sapiens.  
XX FN WO200240715-A2.  
XX PC 23-MAY-2002.  
XX PF 06-SEP-2001; 2501WO-0327628.  
XX PR 06-SEP-2000; 2000US-230505P.  
FR 06-SEP-2000; 2000US-230514P.  
FR 06-SEP-2000; 2000US-230515P.  
PR 06-SEP-2000; 2000US-230517P.  
PR 06-SEP-2000; 2000US-230518P.  
PR 06-SEP-2000; 2000US-230519P.  
PR 06-SEP-2000; 2000US-230597P.  
PR 06-SEP-2000; 2000US-230598P.  
PR 06-SEP-2000; 2000US-230599P.  
PR 06-SEP-2000; 2000US-230610P.  
PR 06-SEP-2000; 2000US-230655P.  
PR 06-SEP-2000; 2000US-230988P.  
PR 06-SEP-2000; 2000US-230989P.  
PR 07-SEP-2000; 2000US-230995P.

PT Nucleic acids encoding polypeptides with cytokine-like activities,  
PT useful in diagnosis and gene therapy -  
XX Claim 1; Page 4838; 6221pp; English.  
XX  
CC The invention relates to polynucleotides (AAK51456-AAK53435) and the  
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666  
CC (AAM80020) are omitted as the relevant pages from the sequence listing  
CC were missing at the time of publication.  
XX  
SQ Sequence 2209 BP; 674 A; 413 C; 544 G; 577 T; 1 other:  
  
Query Match 36.3%; Score 1039.8; DB 22; Length 2209;  
Best Local Similarity 98.1%; Pred. No. 1.3e-259;  
Matches 1107; Conservative 0; Mismatches 12; Indels 10; Gaps 5;  
  
Qy 1747 AAGGGAGAGCAGGATCCCTGGTGTCTCCGGGCGAGATGGAAACACGGGGAAGAAC 63  
Db 4 AAGGGAGAGCAGGATCCCTGGTGTCTCCGGGCGAGATGGAAACACGGGGAAGAAC 63  
  
Qy 1802 CTGGAATCCAGGATTT-CCTGGAAACCGAGATTAATGGG-CMAAAGGGAGAAATG 1859  
Db 64 CTGGAATCCCGAGATTTCTCTGAAACCGAGATTAATGGGCCCCAAAAGGGAGAAATG 123  
  
Qy 1859 GGCT-CCAGGACAGCAAGGAAAGAAAGGAGCCAGGATG-CCTGGTTTAATGGAG 1916  
Db 124 GGCTCCAGGACAGCAAGGAAAGAAAGGAGCCAGGATGCTGGTTTAATGGAG 183  
  
Qy 1917 CAATGGCTCACCAGGCGAGCTGGAAACCGGATCTAAGGGAGCAAGGTTGAACCTGG 1976  
Db 184 CAATGGCTCACCAGGCGAGCTGGAAACCGGATCTAAGGGAGCAAGGTTGAACCTGG 243  
  
Qy 1977 AATTCAGGATGCTGGGCTTCTGGCTCAGGAGAACAGGAGACGGGTTCC 2036  
Db 244 AATTCAGGATGCTGGGCTTCTGGCTCAGGAGAACAGGAGACGGGTTCC 303  
  
Qy 2037 AGGAGAACCGAGATACATGGGTTTACCCGGGATTCAGGAAAGGAGGCAAGAGAA 2096  
Db 304 AGGAGAACCGAGATACATGGGTTTACCCGGGATTCAGGAAAGGAGGCAAGAGAA 363  
  
Qy 2097 TCAAGGTGAAAAGGATTCAGGTTCAAAAGGAGAAATGGAAGCAGGAAATCCAGG 2156  
Db 364 TCAAGGTGAAAAGGATTCAGGTTCAAAAGGAGAAATGGAAGCAGGAAATCCAGG 423  
  
Qy 2157 GCAACAGGAAATTCAGGCTCATGTGCAAAAGGAGAGAGAGGTGAAAGGAGAAC 2216  
Db 424 GCAACAGGAAATTCAGGCTCATGTGCAAAAGGAGAGAGAGGTGAAAGGAGAAC 483  
  
Qy 2217 TGGTGTCCAGGTCCTTGGATCAAAAGGAGAAATCTGGGTCGATGGCTTATGAGGGCC 2276  
Db 484 TGGTGTCCAGGTCCTTGGATCAAAAGGAGAAATCTGGGTCGATGGCTTATGAGGGCC 543  
  
Qy 2277 CGCAGGTCTTAAGGGCACTCGGGATCCAGGTCCTCAGGAGCCCCAGGTTTGGATGG 2336  
Db 544 CGCAGGTCTTAAGGGCACTCGGGATCCAGGTCCTCAGGAGCCCCAGGTTTGGATGG 603  
  
Qy 2337 GAAGCCCGGAAGAGATTTTCAGAACAAATTTATTCGACAAATTTGCACAGATGTAATAG 2396  
Db 604 GAAGCCCGGAAGAGATTTTCAGAACAAATTTATTCGACAAATTTGCACAGATGTAATAG 663  
  
Qy 2397 AGCCAGCTACAGTCTTACTTCAGAGTGGAGAAATAGAAATTTGATCATTCCTGTC 2456  
Db 664 AGCCAGCTACAGTCTTACTTCAGAGTGGAGAAATTAGAAATTTGATCATTCCTGTC 723

Qy 2457 CCAACATGGTCCCGGGTATTCTTGGGCCACCTGGTCCGATAGAGCCAGAGGGTCCGAG 2516  
Db 724 CCAACATGGTCCCGGGTATTCTTGGGCCACCTGGTCCGATAGAGCCAGAGGGTCCGAG 783  
  
Qy 2517 AGGATTACCTGGTTTCCAGGAAAGAGATGGTCTTCTGGATTAGTGGGTGTCCTTGGAGC 2576  
Db 784 AGGATTACCTGGTTTCCAGGAAAGAGATGGTCTTCTGGATTAGTGGGTGTCCTTGGAGC 843  
  
Qy 2577 TCAGGTGTCAGAGGATTTAAAGGCTTACCAGGAAAGAAATGGGAAAAAGGAGCAAGG 2636  
Db 844 TCAGGTGTCAGAGGATTTAAAGGCTTACCAGGAAAGAAATGGGAAAAAGGAGCAAGG 903  
  
Qy 2637 GTTGGGTATCTGGAGAACAGGTCCTCTCGTCCCCCAGGTCAGAGGGCCCTTCTGG 2696  
Db 904 GTTGGGTATCTGGAGAACAGGTCCTCTCGTCCCCCAGGTCAGAGGGCCCTTCTGG 963  
  
Qy 2697 AATAAGCAAGAGGTCCTTCCAGGAGACCCAGGTTCTTCCCTGGCAAGATGGAGCATGG 2756  
Db 964 AATAAGCAAGAGGTCCTTCCAGGAGACCCAGGTTCTTCCCTGGCAAGATGGAGCATGG 1023  
  
Qy 2757 ABAACCTGGATCCNAGGCGCAACCGGCCCCCGAGGCATCTCGACCCATCATGTATT 2816  
Db 1024 ABAACCTGGATCCNAGGCGCAACCGGCCCCCGAGGCATCTCGACCCATCATGTATT 1083  
  
Qy 2817 TAGTGTAATTGCCAGAGAGATCCGTTTCAGAAAAGGAGCAAACTATTAG 2865  
Db 1084 TAGTGTAATTGCCAGAGAGATCCGTTTCAGAAAAGGAGCAAACTATTAG 1132  
  
RESULT 11  
AA:60188  
ID AAT60188.o:standar.d; cDNA; 2209 BP.  
XX AC AA:60188;  
XX  
XX  
DT 22-OCT-2001 (first entry)  
XX  
DE Human polynucleotide SEQ ID NO 4177.  
XX  
KW Human; nontopic; immunosuppressant; cytostatic; gene therapy; cancer;  
KW peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
KW leukaemia; ss.  
XX  
OS Homo sapiens.  
XX  
XX  
PN WO200153312-A1.  
XX  
PD 26-JUL-2001.  
XX  
PF 26-DEC-2000; 2000WO-US34263.  
XX  
PR 21-JAN-2000; 2000US-0488725.  
PR 25-APR-2000; 2000US-0552117.  
PR 09-JUL-2000; 2000US-0598042.  
PR 19-JUL-2000; 2000US-0620312.  
PR 03-AUG-2000; 2000US-0653450.  
PR 14-SEP-2000; 2000US-0662191.  
PR 19-OCT-2000; 2000US-0693036.  
PR 29-NOV-2000; 2000US-0727344.  
XX  
XX (HYSB-) HYSEQ INC.  
PA  
XX  
PI Tang Yt, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
PI Zhao QA, Zhou P, Goodrich R, Drmanac R;  
XX  
XX WPI; 2001-442253/47.  
DR  
DR P-PSDB; XAN41032.  
XX  
XX



XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;  
 PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
 PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;  
 XX  
 DR WPI; 2001-476283/51.  
 DR P-PSDB; AAM78912.  
 XX  
 PT Nucleic acids encoding polypeptides with cytokine-like activities,  
 PT useful in diagnosis and gene therapy -  
 XX  
 PS Claim 1; Page 2132-2134; 6221pp; English.  
 XX  
 CC The invention relates to polynucleotides (AAK51456-AAK53435) and the  
 CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activin/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation.  
 CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666  
 CC (AAM80020) are omitted as the relevant pages from the sequence listing  
 CC were missing at the time of publication.  
 XX  
 SQ Sequence 2230 BP; 679 A; 409 C; 550 G; 592 T; 0 other;

Query Match 37.1%; Score 1062; DB 22; Length 2230;  
 Best Local Similarity 100.0%; Pred. No. 2.1e-265;  
 Matches 1062; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1804 GGAATCCAGGATTTCTCTGMAACCCAGGATTAATGGGCCAAAGGGAGAAATGGGGCT 1863  
 Db 91 GGANTCCAGGATTTCTGGNAACCGGANTTAATGGGCCAAAGGGAGAAATGGGGCT 150  
 Qy 1864 CCAGGACAGCAAGAAAAGAGCCCGAGGATCCCTGGTTTAAATGGGAAGCAATGGC 1923  
 Db 151 CCAGGACAGCAAGAAAAGAGCCCGAGGATCCCTGGTTTAAATGGGAAGCAATGGC 210  
 Qy 1924 TCACCGAGCCAGCTGACACCGGATCTAGGGAGCAAGGTCGACCTGGATTCNA 1983  
 Db 211 TCACCGAGCCAGCTGACACCGGATCTAGGGAGCAAGGTCGACCTGGATTCNA 270  
 Qy 1984 GGGATGCTGGGCTTCTGGGCTCAAGGGAGAACACAGGACACGGTTCCCGAGAGAA 2043  
 Db 271 GGGATGCTGGGCTTCTGGGCTCAAGGGAGAACACAGGACACGGTTCCCGAGAGAA 330  
 Qy 2044 CCAGGATACATGGGTTTACCCGGATTCAGGGAAGAAAGGGGACAAAGGAATCAAGT 2103  
 Db 331 CCAGGATACATGGGTTTACCCGGATTCAGGGAAGAAAGGGGACAAAGGAATCAAGT 390  
 Qy 2104 GAAAGAGGATTCAGGGTCAAAAGGGAGAAATGGAAGACAGGGAATTCAGGGCAACAG 2163  
 Db 391 GAAAGAGGATTCAGGGTCAAAAGGGAGAAATGGAAGACAGGGAATTCAGGGCAACAG 450  
 Qy 2164 GGAATTCAGGCCATCATGTGTCAAAAGGAGAGAGGTGAAAGGGAGAACCTGGTGTC 2223  
 Db 451 GGAATTCAGGCCATCATGTGTCAAAAGGAGAGAGGTGAAAGGGAGAACCTGGTGTC 510  
 Qy 2224 CGAGGTGCCATTTGATCAAAAGGAGAAATCTGGGGTGGATGGCTTGTATGGGGCCCGCAGT 2283  
 Db 511 CGAGGTGCCATTTGATCAAAAGGAGAAATCTGGGGTGGATGGCTTGTATGGGGCCCGCAGT 570  
 Qy 2284 CCTAAGGGGCAACCTGGGGATTCAGGTCCTCAGGACCCCGCAGGTTTGGATGGGAAGGCC 2343  
 Db 571 CCTAAGGGGCAACCTGGGGATTCAGGTCCTCAGGACCCCGCAGGTTTGGATGGGAAGGCC 630  
 Qy 2344 GGAAGAGGTTTTTCAGAACAAATTTATTCGACAAAGTTTGCACAGATGTAATAGAGCCAG 2403  
 Db 631 GGAAGAGGTTTTTCAGAACAAATTTATTCGACAAAGTTTGCACAGATGTAATAGAGCCAG 690

Qy 2404 CTACCACTCTTACTTCAGAGTGGAGAAATTAGAAATTTGTGATCTATGCTCTCCACAT 2463  
 Db 691 CTACCACTCTTACTTCAGAGTGGAGAAATTAGAAATTTGTGATCTATGCTCTCCACAT 750  
 Qy 2464 GGCTCCCCGGGTATTCTCTGGGCCACCTGGTCCGATAGGCCAGAGGGTCCAGAGGATTA 2523  
 Db 751 GGCTCCCCGGGTATTCTCTGGGCCACCTGGTCCGATAGGCCAGAGGGTCCAGAGGATTA 810  
 Qy 2524 CTTGGTTTCCAGGAGAGATGGTCTTCTGGATTTAGTGGGTCTCCCTGGAGCTCCAGGT 2583  
 Db 811 CTTGGTTTCCAGGAGAGATGGTCTTCTGGATTTAGTGGGTCTCCCTGGAGCTCCAGGT 870  
 Qy 2584 GTCCAGAGGATTAAGAGGCTTACACAGAAAGAAATGGGAAAGAGGAGCCAGGGTTTGGG 2643  
 Db 871 GTCCAGAGGATTAAGAGGCTTACACAGAAAGAAATGGGAAAGAGGAGCCAGGGTTTGGG 930  
 Qy 2644 TATCTGGAGAACAAAGGTCTCTGGTCTCCCGAGTCCAGAGGGCCCTCTTGGAAATAGC 2703  
 Db 931 TATCTGGAGAACAAAGGTCTCTGGTCTCCCGAGTCCAGAGGGCCCTCTTGGAAATAGC 990  
 Qy 2704 AAGAAAGGTCTCTCCAGGAGACCCAGGTCTCTCTGGCAAGATGGAGACCATGGAACCT 2763  
 Db 991 AAGAAAGGTCTCTCCAGGAGACCCAGGTCTCTCTGGCAAGATGGAGACCATGGAACCT 1050  
 Qy 2764 GGAATCCAAAGGGCAACCCAGGCCCGCCAGGCATCTCCGACCCCATCATCTATGTTTGTGTA 2823  
 Db 1051 GGAATCCAAAGGGCAACCCAGGCCCGCCAGGCATCTCCGACCCCATCATCTATGTTTGTGTA 1110  
 Qy 2824 ATTGCCAGAGAGATCCGTTTCAGAAAGGACCAAACTATTAG 2865  
 Db 1111 ATTGCCAGAGAGATCCGTTTCAGAAAGGACCAAACTATTAG 1152

RESULT 9  
 AAI58402  
 ID AAI58402 standard; cDNA; 2230 BP.  
 XX  
 AC AAI58402;  
 XX  
 DT 22-OCT-2001 (first entry)  
 XX  
 DE Human polynucleotide SEQ ID NO 605.  
 XX  
 KW Human; neotropic; immunosuppressant; cytostatic; gene therapy; cancer;  
 KW peripheral nervous system; neuropathy; central nervous system; CNS;  
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
 KW leukaemia; ss.  
 KW  
 OS Homo sapiens.  
 XX  
 FN WO200153312-A1.  
 XX  
 PD 26-JUL-2001.  
 XX  
 PF 26-DEC-2000; 2000WO-US34263.  
 XX  
 PR 21-JAN-2000; 2000US-0488725.  
 PR 25-APR-2000; 2000US-0552317.  
 PR 09-JUL-2000; 2000US-0598042.  
 PR 19-JUL-2000; 2000US-0620312.  
 PR 03-AUG-2000; 2000US-0653450.  
 PR 14-SEP-2000; 2000US-0662191.  
 PR 19-OCT-2000; 2000US-0693036.  
 PR 29-NOV-2000; 2000US-0727344.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX The invention relates to human novel polypeptides and their associated  
 CC polynucleotides. The polypeptides and polynucleotides are useful in gene  
 CC therapy for treating, inhibiting or preventing neural disorders, immune  
 CC system disorders (e.g. systemic lupus erythematosus, rheumatoid  
 CC arthritis and multiple sclerosis), muscular disorders, respiratory  
 CC diseases (e.g. nasal vestibulitis, nasal polyps and sinusitis),  
 CC reproductive disorders, gastrointestinal disorders, pulmonary disorders,  
 CC cardiovascular disorders (e.g. congenital heart defects, Einstein's  
 CC anomaly and hypoplastic left heart syndrome), renal disorders (e.g. acute  
 CC kidney failure and end-stage renal disease), hyperproliferative disorders  
 CC (e.g. Hodgkin's disease and leukaemia), inflammatory diseases (e.g.  
 CC septic shock, bursitis and appendicitis), allergic reactions and  
 CC conditions (e.g. asthma), blood related disorders (e.g. thrombosis,  
 CC atherosclerosis and myocardial infarction) and cancerous diseases.  
 CC Sequences ABX73173-ABX74167 represent human novel polynucleotides of the  
 CC invention.  
 XX  
 SQ Sequence 2375 BP; 731 A; 435 C; 600 G; 609 T; 0 other;  
 Query Match 44.7%; Score 1282; DB 25; Length 2375;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1282; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1584 AAAGGCTGAATGGTGGTCCAAAGGAGCAAAAGGATCACCTGGATTTTATGGCAAAAGGG 1643  
 DB 9 AAAGGCTGAATGGTGGTCCAAAGGAGCAAAAGGATCACCTGGATTTTATGGCAAAAGGG 68  
 QY 1644 TGCAAAAGGTGAAGAGGGAGTGGTGGTCCCTGGCTCCCTGGAGCTGGTGGAGAAC 1703  
 DB 69 TGCAAAAGGTGAAGAGGGAGTGGTGGTCCCTGGCTCCCTGGAGCTGGTGGAGAAC 128  
 QY 1704 AGGAAGACATGGAAGGATGGAATTAATGGGTAGTCCGGTTTCAAGGGAGAACGAGGATC 1763  
 DB 129 AGGAAGACATGGAAGGATGGAATTAATGGGTAGTCCGGTTTCAAGGGAGAACGAGGATC 188  
 QY 1764 CCCTGGTGTCTCCGGGAGGATGGAACACGGGGAGGCTGGATCCAGGATTTCTGG 1823  
 DB 189 CCCTGGTGTCTCCGGGAGGATGGAACACGGGGAGGCTGGATCCAGGATTTCTGG 248  
 QY 1824 AAACCGAGGATTAATGGGCCAAAAGGAGCAAAATTTGGGCTCCAGGACAGCAAGGAAAAA 1883  
 DB 249 AAACCGAGGATTAATGGGCCAAAAGGAGCAAAATTTGGGCTCCAGGACAGCAAGGAAAAA 308  
 QY 1884 AGGAGCCCCAGGATGCTGGTTTAATGGGAGCAATGGCTCACCAGGCGAGCTGGAAC 1943  
 DB 309 AGGAGCCCCAGGATGCTGGTTTAATGGGAGCAATGGCTCACCAGGCGAGCTGGAAC 368  
 QY 1944 ACCGGGATCTAAGGAGCAAGAGTGAACCTGGAAATTCAGGGATGCTGGGCTTCTGG 2003  
 DB 369 ACCGGGATCTAAGGAGCAAGAGTGAACCTGGAAATTCAGGGATGCTGGGCTTCTGG 428  
 QY 2004 GCTCAAGGGAGAACAGGAGCAAGGTTTCCCGAGGAGAACAGGATACATGGGTTTACC 2063  
 DB 429 GCTCAAGGGAGAACAGGAGCAAGGTTTCCCGAGGAGAACAGGATACATGGGTTTACC 488  
 QY 2064 CGGGATTCAGGAGAAAAAGGGGACAAAGGAAATCAAGGTCAAAAGGATTCAGGGTCA 2123  
 DB 489 CGGGATTCAGGAGAAAAAGGGGAGCAAAAGGAAATCAAGGTCAAAAGGATTCAGGGTCA 548  
 QY 2124 AAAGGAGAAATGGAAGACAGGGAAATTCAGGGCAACAGGGAAATTCAGGGCCATCATGG 2183  
 DB 549 AAAGGAGAAATGGAAGACAGGGAAATTCAGGGCAACAGGGAAATTCAGGGCCATCATGG 608  
 QY 2184 TGCAAAAGGAGAGAGGTGAAGAGGAGAACCTGGTCCGAGGTGCCATTCGATCAAA 2243  
 DB 609 TGCAAAAGGAGAGAGGTGAAGAGGAGAACCTGGTCCGAGGTGCCATTCGATCAAA 668  
 QY 2244 AGGAGAAATCTGGGGTGGATGGCTTGAATGGGGCCCGCAGGCTCAAGGGGCAAACTGGGGA 2303  
 DB 669 AGGAGAAATCTGGGGTGGATGGCTTGAATGGGGCCCGCAGGCTCAAGGGGCAAACTGGGGA 728  
 QY 2304 TCCAGGTCTCAGGACCCCGCAGGTTCGATGGGAAGCCCGGAGAGAGTTTCAGAAC 2363

DB 729 TCCAGGTCTCCAGGGACCCCGAGTTTGGATGGAGAGCCCGGAGAGAGATTTCASACA 788  
 QY 2364 ATTATTTCGACAAAGTTTGGCAGAGATGTAATAAGAGCCCGAGCTACCAAGTCTTACTTCAGAG 2423  
 DB 789 ATTATTTCGACAAAGTTTGGCAGAGATGTAATAAGAGCCCGAGCTACCAAGTCTTACTTCAGAG 849  
 QY 2424 TGAAGAAATTAGAAATTTGATCATTCGCTGTCGCAACATGGCTCCCGGGTATTCCTGG 2483  
 DB 849 TGAAGAAATTAGAAATTTGATCATTCGCTGTCGCAACATGGCTCCCGGGTATTCCTGG 908  
 QY 2484 GCCACCTGGTCCGATAGGCCCCAGAGGGTCCCGAGAGATTAACCTGGTTTCCAGGAGAGAGA 2543  
 DB 909 GCCACCTGGTCCGATAGGCCCCAGAGGGTCCCGAGAGATTAACCTGGTTTCCAGGAGAGAGA 968  
 QY 2544 TGGTGTCTCTGGATTTAGTGGGTGCTCCCTGGAAGTCCAGGTGTCAGAGATTTAAAGGGCT 2603  
 DB 969 TGGTGTCTCTGGATTTAGTGGGTGCTCCCTGGAAGTCCAGGTGTCAGAGATTTAAAGGGCT 1028  
 QY 2604 ACCAGGAAGAAATGGGGAAAAAGGAGCCAAAGGGTTTGGGTATCTCTGGAGAACAAAGGTCC 2663  
 DB 1029 ACCAGGAAGAAATGGGGAAAAAGGAGCCAAAGGGTTTGGGTATCTCTGGAGAACAAAGGTCC 1088  
 QY 2664 TCCTGGTCCCCAGGTCCAGAGGGCCCTCTCTGGAATAAGCAAGAGGTCTCTCCAGGAGA 2723  
 DB 1089 TCCTGGTCCCCAGGTCCAGAGGGCCCTCTCTGGAATAAGCAAGAGGTCTCTCCAGGAGA 1148  
 QY 2724 CCAGGTCTCCCTGGCAAGATGGAGACCATGGAACCTGGAATCCAAAGGSCCAACCGG 2783  
 DB 1149 CCAGGTCTCCCTGGCAAGATGGAGACCATGGAACCTGGAATCCAAAGGSCCAACCGG 2808  
 QY 2784 CCCTCCAGGATCTGGAGCCATCATATGTTTTAGTGTAAATTCGAGAGAGATCCGTT 2843  
 DB 1209 CCCTCCAGGATCTGGAGCCATCATATGTTTTAGTGTAAATTCGAGAGAGATCCGTT 1268  
 QY 2844 CAGAAAGGACCAACCTATTAG 2865  
 DB 1269 CAGAAAGGACCAACCTATTAG 1290  
 RESULT 8  
 AAK52045  
 ID AAK52045 standard; cDNA; 2230 BP.  
 AC AAK52045;  
 XX  
 D7 06-NOV-2001 (first entry)  
 XX Human polynucleotide SEQ ID NO 590.  
 DE Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
 KW nervous system disorder; arthritis; inflammation; ss.  
 XX Homo sapiens.  
 OS  
 XX  
 PN WO2001:57190-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 05-FEB-2001; 2001WO-US04098.  
 XX  
 PR 03-FEB-2000; 2000US-0496914.  
 PR 27-APR-2000; 2000US-0560875.  
 PR 23-JUN-2000; 2000US-0598075.  
 PR 19-JUL-2000; 2000US-0620325.  
 PR 02-SEP-2000; 2000US-0654936.  
 PR 15-SEP-2000; 2000US-0663561.  
 PR 20-OCT-2000; 2000US-0693325.  
 PR 30-NOV-2000; 2000US-0728422.  
 XX  
 PA (HYSE-) HYSEQ INC.

Qy 2124 AAAGGGAAGAAATGGAGACAGGGAATTCAGGGCAACAGGGAATTCAGGCCATCATGG 2183  
 Db |||||||  
 Qy 549 AAAGGGAAGAAATGGAGACAGGGAATTCAGGGCAACAGGGAATTCAGGCCATCATGG 608  
 Db |||||||  
 Qy 2184 TGCAAAAGGAGAGAGAGGTGAAAGGAGAGAACCTGGTGCCGAGGTGCCATTGGATCAAA 2243  
 Db |||||||  
 Qy 609 TGCAAAAGGAGAGAGAGGTGAAAGGAGAGAACCTGGTGCCGAGGTGCCATTGGATCAAA 668  
 Db |||||||  
 Qy 2244 AGGAGAACTGGGGTGGATGGCTTGATGGGGCCGAGGTCTTAAGGGGCAACCTGGGGA 2303  
 Db |||||||  
 Qy 669 AGGAGAACTGGGGTGGATGGCTTGATGGGGCCGAGGTCTTAAGGGGCAACCTGGGGA 728  
 Db |||||||  
 Qy 2304 TCCAGGTCTCTCAGGAGCCCCCAGGTTTGGATGGGAAGCCCGGAAGAGAGTTTTCAGAAC 2363  
 Db |||||||  
 Qy 729 TCCAGGTCTCTCAGGAGCCCCCAGGTTTGGATGGGAAGCCCGGAAGAGAGTTTTCAGAAC 788  
 Db |||||||  
 Qy 2364 ATTATTTCGACAAAGTTTGCACAGATGTAATAGAGGCCAGCTACCAAGTCTTACTTCAGAG 2423  
 Db |||||||  
 Qy 789 ATTATTTCGACAAAGTTTGCACAGATGTAATAGAGGCCAGCTACCAAGTCTTACTTCAGAG 848  
 Db |||||||  
 Qy 2424 TGAAGAAATTAGAAATTTGATCATTTGCTGTCCTCCCAACATGGCTCCCGGGTATTCCTGG 2483  
 Db |||||||  
 Qy 849 TGGAGAAATTAGAAATTTGATCATTTGCTGTCCTCCCAACATGGCTCCCGGGTATTCCTGG 908  
 Db |||||||  
 Qy 2484 GCCACCTGGTCCGATAGGCCAGAGGGTCCAGAGGATTACCTGGTTTCCAGGAGAGA 2543  
 Db |||||||  
 Qy 909 GCCACCTGGTCCGATAGGCCAGAGGGTCCAGAGGATTACCTGGTTTCCAGGAGAGA 968  
 Db |||||||  
 Qy 2544 TGGTGTCTCTGGATTTAGTGGGTGTCCTGACCTCCAGGTGTCAGAGGATTAAAGGCTT 2603  
 Db |||||||  
 Qy 969 TGGTGTCTCTGGATTTAGTGGGTGTCCTGACCTCCAGGTGTCAGAGGATTAAAGGCTT 1028  
 Db |||||||  
 Qy 2604 ACCAGGAAGAAATGGGCAAAAGGAGCCCAAGGGTTTGGGTATCCTGGAGAACCAAGGTCC 2663  
 Db |||||||  
 Qy 1029 ACCAGGAAGAAATGGGCAAAAGGAGCCCAAGGGTTTGGGTATCCTGGAGAACCAAGGTCC 1088  
 Db |||||||  
 Qy 2664 TCTGTGTCCTCCAGGTCCAGAGGCCCTCTCGAATAGCAAAAGAGGTCTCCAGGAGA 2723  
 Db |||||||  
 Qy 1089 TCTGTGTCCTCCAGGTCCAGAGGCCCTCTCGAATAGCAAAAGAGGTCTCCAGGAGA 1148  
 Db |||||||  
 Qy 2724 CCCAGGTCTCTCTGGCAAGATGGAGACCATGAAACCTGGAATCAAGGGCAACCAAG 2783  
 Db |||||||  
 Qy 1149 CCCAGGTCTCTCTGGCAAGATGGAGACCATGAAACCTGGAATCAAGGGCAACCAAG 1208  
 Db |||||||  
 Qy 2784 CCCCCAGGATCTGGACCCATCATATGTTTGTATTTAGTGAATTTGCGAGAGATCCGTT 2843  
 Db |||||||  
 Qy 1209 CCCCCAGGATCTGGACCCATCATATGTTTGTATTTAGTGAATTTGCGAGAGATCCGTT 1269  
 Db |||||||  
 Qy 2844 CAGAAAGGACCAAACTATTAG 2865  
 Db |||||||  
 Qy 1269 CAGAAAGGACCAAACTATTAG 1290  
 Db |||||||

RESULT 7  
 ABX73394  
 ID ABX73394 standard; DNA; 2375 BP.

AC ABX73394;  
 DT 18-MAR-2003 (first entry)  
 DE Human novel polynucleotide #222.

KW Human; gene; ds; neural disorder; immune system disorder; renal disorder;  
 KW muscular disorder; respiratory disease; reproductive disorder;  
 KW gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;  
 KW hyperproliferative disorder; inflammatory disease; allergic reaction;  
 KW blood related disorder; cancer; immunosuppressive; antineoplastic;  
 KW cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;  
 KW haemostatic; antiarteriosclerotic.  
 XX Homo sapiens.

XX US2002132753-A1.  
 XX 19-SEP-2002.  
 XX 17-JAN-2001; 2001US-0764864.  
 XX 31-JAN-2000; 2000US-179065P.  
 XX 04-FEB-2000; 2000US-180628P.  
 XX 28-JUN-2000; 2000US-214886P.  
 XX 07-JUL-2000; 2000US-216647P.  
 XX 07-JUL-2000; 2000US-216883P.  
 XX 11-JUL-2000; 2000US-217487P.  
 XX 11-JUL-2000; 2000US-217496P.  
 XX 14-JUL-2000; 2000US-218290P.  
 XX 26-JUL-2000; 2000US-220963P.  
 XX 26-JUL-2000; 2000US-220964P.  
 XX 14-AUG-2000; 2000US-224518P.  
 XX 14-AUG-2000; 2000US-224519P.  
 XX 14-AUG-2000; 2000US-225267P.  
 XX 14-AUG-2000; 2000US-225268P.  
 XX 14-AUG-2000; 2000US-225270P.  
 XX 14-AUG-2000; 2000US-225447P.  
 XX 14-AUG-2000; 2000US-225477P.  
 XX 14-AUG-2000; 2000US-225757P.  
 XX 14-AUG-2000; 2000US-225758P.  
 XX 22-AUG-2000; 2000US-226868P.  
 XX 30-AUG-2000; 2000US-228924P.  
 XX 01-SEP-2000; 2000US-229287P.  
 XX 01-SEP-2000; 2000US-229343P.  
 XX 01-SEP-2000; 2000US-229344P.  
 XX 01-SEP-2000; 2000US-229345P.  
 XX 05-SEP-2000; 2000US-229509P.  
 XX 05-SEP-2000; 2000US-229513P.  
 XX 08-SEP-2000; 2000US-231413P.  
 XX 21-SEP-2000; 2000US-234223P.  
 XX 21-SEP-2000; 2000US-234274P.  
 XX 25-SEP-2000; 2000US-234977P.  
 XX 27-SEP-2000; 2000US-235834P.  
 XX 29-SEP-2000; 2000US-236327P.  
 XX 29-SEP-2000; 2000US-236367P.  
 XX 29-SEP-2000; 2000US-236368P.  
 XX 29-SEP-2000; 2000US-236369P.  
 XX 29-SEP-2000; 2000US-236370P.  
 XX 02-OCT-2000; 2000US-236802P.  
 XX 02-OCT-2000; 2000US-237037P.  
 XX 02-OCT-2000; 2000US-237038P.  
 XX 02-OCT-2000; 2000US-237039P.  
 XX 02-OCT-2000; 2000US-237040P.  
 XX 13-OCT-2000; 2000US-239935P.  
 XX 20-OCT-2000; 2000US-240960P.  
 XX 20-OCT-2000; 2000US-241785P.  
 XX 20-OCT-2000; 2000US-241809P.  
 XX 01-NOV-2000; 2000US-244617P.  
 XX 17-NOV-2000; 2000US-249299P.  
 XX 08-DEC-2000; 2000US-251856P.  
 XX 08-DEC-2000; 2000US-251868P.  
 XX 08-DEC-2000; 2000US-251869P.  
 XX (ROSE/) ROSEN C A.  
 XX (RUBE/) RUBEN S M.  
 XX (BARA/) BARASH S C.

PI Rosen CA, Ruben SM, Barash SC;

WPI; 2003-147444/14.  
 P-PSDB; ABUS5134.

New polypeptides and nucleic acids, useful in gene therapy for  
 treating, inhibiting or preventing e.g. neural immune system,  
 muscular, respiratory, reproductive, gastrointestinal, pulmonary,  
 cardiovascular or renal disorders -

Claim 1; SEQ ID NO 232; 402pp; English.



QY	901	TTAACTATTGATGGAAGGCCAACAAATAGCAGTTACCTTAAATGGTGTGGACAAATCTTA	960	XX
DB	1015	TTAACTATTGATGG-ATGCCAACAATAGCAGTTACCTTAAATGGTGTGGACAAATCTTA	1073	PD
QY	961	TTATTTACAAACCGCGCTAATTAATGGCTCACAAGTGGTTACCTTTGCTTAACCTCAA	1020	XX
DB	1074	TTATTTACAAACCGCGCTAATTAATGGCTCACAAGTGGTTACCTTTGCTTAACCTCAA	1133	PF
QY	1021	GTTAAGACCTGTTTGTATGAAGCTGGCACAATTCGTCTCTTAGTACAGCAAGAT	1080	XX
DB	1134	GTTAAGACCTGTTTGTATGAAGCTGGCACAATTCGTCTCTTAGTACAGCAAGAT	1193	PR
QY	1081	GTGACTTTGTATATTGATGACCAACAAATTTGAAAAAAGCCCTTACATCCAGTTTAGGG	1140	PR
DB	1194	G-GACTTTGTATATTGATGACCAACAAATTTGAAAAAAGCCCTTACATCCAGTTTAGGG	1253	PR
QY	1141	ATCTTGATCAATTTGGGCAACCCCAATTTGAAAAAATCTTGGAAAAAGAAACTGTTTCA	1200	PR
DB	1254	ATCTTGATCAATTTGGGCAACCCCAATTTGAAAAAATCTTGGAAAAAGAAACTGTTTCA	1313	PR
QY	1201	TTTGATGTCACAAAGTTGGAACTTACTGTGACCCAGAACACAAACCGGGAGACACA	1260	PR
DB	1314	TTTGATGTCACAAAGTTGGAACTTACTGTGACCCAGAACACAAACCGGGAGACACA	1373	PR
QY	1261	TCTGATTCCTGGATTTTTCCTTAATGGTCCAGTGATGTAGGTTCAACTCCAGTCCC	1320	PR
DB	1374	TCTGATTCCTGGATTTTTCCTTAATGGTCCAGTGATGTAGGTTCAACTCCAGTCCC	1433	PR
QY	1321	TGTATTTGCTCCCGGAAACCGGACTTCAAGCCCCCAAGGTGACCTGGACTGCT	1380	PR
DB	1434	TGTATTTGCTCCCGGAAACCGGACTTCAAGCCCCCAAGGTGACCTGGACTGCT	1493	PR
QY	1381	GGGAACCTCGGTACCTCGGACAACTGCTCAAGATGGTAAGCCTT-----	1425	PR
DB	1494	GGGAACCTCGGTACCTCGGACAACTGCTCAAGATGGTAAGCCTT-----	1553	PR
QY	1426	-----GGATATCAGGGAATTCAGGGACACAGGT	1455	PR
DB	1554	TTAGTCATCTCCGGTATATCTCGGATTACAGGATATCAGGGAATTCAGGGACACAGGT	1613	PR
QY	1456	GTTCCAGGATCTCCAGGAATACAGGAGTCCAGGACTCAGGACTTCAAAGGAGAACCA	1515	PR
DB	1614	GTTCCAGGATCTCCAGGAATACAGGAGTCCAGGACTCAGGACTTCAAAGGAGAACCA	1673	PR
QY	1516	GGCGGAGATGGTGACAAG	1533	PR
DB	1674	GGCGGAGATGGTGACAAG	1691	PR
RESULT 6				XX
AAS26053				XX
ID AAS26053 standard; cDNA; 2375 BP.				XX
AC AAS26053;				XX
DT 07-NOV-2001 (first entry)				XX
DE Human cDNA encoding a novel secreted protein. Seq ID 232.				XX
KW Human; immunosuppressive; antiarthritic; ss; antirheumatic;				XX
KW cytotoxic; cardiant; vasotropic; cerebroprotective; neurotropic;				XX
KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;				XX
KW vulnary; secreted protein; rheumatoid arthritis;				XX
KW hyperproliferative disorder; cardiovascular disorder; cardiac arrest;				XX
KW cerebrovascular disorder; cerebral ischemia; angiogenesis;				XX
KW nervous system disorder; Alzheimer's disease; infection; ocular disorder;				XX
KW corneal infection; wound healing; epithelial cell proliferative;				XX
KW skin aging; food additive; preservative; antiproliferative.				XX
OS Homo sapiens.				OS
XX WO200155322-A2.				XX
PN				PN



```

Db 1321 TGTATTGCTCTCGGAAACACGAGCTTCAAGGCCCAAGGTGACCTGGACTGGCT 1380
Qy 1381 GGAACCTGGCTTACCTCGACCACTGGTCAAGATGGTAAGCCT----- 1425
Db 1381 GGAACCTGGCTTACCTCGACCACTGGTCAAGATGGTAAGCCTGGTACTGAAAGC 1440
Qy 1426 -----GGATATCAGGGAATTCAGGGACACAGGT 1455
Db 1441 TTAGTCATCTCCGGTATATCTCGGATTACAGGATATCAGGGAATTCAGGGACACAGGT 1500
Qy 1456 GTTCCAGGATCTCCAGGAATACAGGAGCTCGAGGACTACAGGTTACAAAGAGAACCA 1515
Db 1501 GTTCCAGGATCTCCAGGAATACAGGAGCTCGAGGACTACAGGTTACAAAGAGAACCA 1560
Qy 1516 GGGGAGATGTCACAG 1533
Db 1561 GGCGAGATGGTACAAG 1578

RESULT 5
AAK52265
ID AAK52265 standard; cDNA; 1691 BP.
XX AC AAK52265;
XX DT 06-NOV-2001 (first entry)
XX DE Human polynucleotide SEQ ID NO 810.
XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX KW tissue growth factor; immunomodulatory; cancer; leukaemia;
XX KW nervous system disorder; arthritis; inflammation; ss.
XX OS Homo sapiens.
XX PN W0200157190-A2.
XX PD 09-AUG-2001.
XX PF 05-FEB-2001; 2001WO-US04098.
XX PR 03-FEB-2000; 2000US-0496914.
XX PR 27-APR-2000; 2000US-0560875.
XX PR 20-JUN-2000; 2000US-0598075.
XX PR 19-JUL-2000; 2000US-0620325.
XX PR 01-SEP-2000; 2000US-0654936.
XX PR 15-SEP-2000; 2000US-0663561.
XX PR 20-OCT-2000; 2000US-0693325.
XX PR 30-NOV-2000; 2000US-0728422.
XX PA (HYSE-) HYSEQ INC.
XX PI Tang Y, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Xia Y;
PI Zhao QA, Wang J, Zhang J, Ren F, Chen R, Wang Z;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX DR WPI; 2001-476283/51.
XX DR P-PSDB; AAM79132.
XX XX Nucleic acids encoding polypeptides with cytokine-like activities,
XX PT useful in diagnosis and gene therapy -
XX PS Claim 1; Page 2710-2711; 6221pp; English.
XX CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
XX CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
XX CC cytokine, cell proliferation or cell differentiation or which may induce
XX CC production of other cytokines in other cell populations. The
XX CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX CC peptide therapy. The polypeptides have various cytokine-like activities,
XX CC e.g. stem cell growth factor activity, haematopoiesis regulating

```

```

CC activity, tissue growth factor activity, immunomodulatory activity and
CC activity/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX SQ Sequence 1691 BP; 532 A; 326 C; 386 G; 447 T; C other;

Query Match 51.1%; Score 1462.8; DB 22; Length 1691;
Best Local Similarity 97.0%; Pred. No. 0;
Matches 1530; Conservative 0; Mismatches 2; Indels 46; Gaps 2;

Qy 1 ATGGCTCACTATATTACATTTCTCTGCATGGTTTGGTGTCTCTTCTTCAGAAATCTCTGG 60
Db 115 ATGGCTCACTATATTACATTTCTCTGCATGGTTTGGTGTCTCTTCTTCAGAAATCTCTGG 174
Qy 61 TTAGCTGAAGATGGGGAAGTAAGATCAAGTTCTGCTACTGCTCCGACAGATTTAGTTTC 120
Db 175 TTAGCTGAAGATGGGGAAGTAAGATCAAGTTCTGCTACTGCTCCGACAGATTTAGTTTC 234
Qy 121 ATCTTAGATGGCTTTATAGTGTGGCCGAGAAACTTTGAAATAGTGCAAAAGTGGCTT 180
Db 235 ATCTTAGATGGCTTTATAGTGTGGCCGAGAAACTTTGAAATAGTGCAAAAGTGGCTT 294
Qy 181 GTCAATATCAAAAAAATTTGACATAGGGCCGAAAGTTATTCAAGTTGGAGTGGTCAA 240
Db 295 GTCAATATCAAAAAAATTTGACATAGGGCCGAAAGTTATTCAAGTTGGAGTGGTCAA 354
Qy 241 TATAGTACTACCTGTGCTGGAGATTCCTTCGGAAGCTNTGATTCAGGAGAACATTTC 300
Db 355 TATAGTACTACCTGTGCTGGAGATTCCTTCGGAAGCTATGATTCAGGAGAACATTTC 414
Qy 301 ACCGCAGCAGTGAATCCATCTCTACTTAGCAGAAACACAAACACAGGGAGGCCATC 360
Db 415 ACCGCAGCAGTGAATCCATCTCTACTTAGCAGAAACACAAACACAGGGAGGCCATC 474
Qy 361 CAGTTTCCGCTCGATTACCTTTTGGCCAAAGTCTCTCAGCATTTCTGACTAAGATAGCAGT 420
Db 474 CAGTTTCCGCTCGATTACCTTTTGGCCAAAGTCTCTCAGCATTTCTGACTAAGATAGCAGT 534
Qy 421 GTACTTACGGATGGGCAATCCGAGATGACGTCAAGATGAGCTCAAGCAGCAGAGAT 480
Db 535 GTACTTACGGATGGGCAATCCGAGATGACGTCAAGATGAGCTCAAGCAGCAGAGAT 594
Qy 481 AGTAAGATAACATTATTGCTATTTGGTGTGCTTCAGAAACAGAAAGATGCCGAACTTAGA 540
Db 595 AGTAAGATAACATTATTGCTATTTGGTGTGCTTCAGAAACAGAAAGATGCCGAACTTAGA 654
Qy 541 GCTATTGCCAACAGCCTTCTGCTACTTATGTGTTTTTATGTTGGAAGACTATATTCGAATA 600
Db 655 GCTATTGCCAACAGCCTTCTGCTACTTATGTGTTTTTATGTTGGAAGACTATATTCGAATA 714
Qy 601 TCCAAATAGGGAAGTGAAGAGCAAGTAACCTTTCTGAGAGATCTGCTGCCAACACAGA 660
Db 715 TCCAAATAGGGAAGTGAAGAGCAAGTAACCTTTCTGAGAGATCTGCTGCCAACACAGA 774
Qy 661 ATTCCAGTGGCAGCTCTGATGAAGGGGATTTGATTTCTTTTAGGTTAGATGTAAT 720
Db 775 ATTCCAGTGGCAGCTCTGATGAAGGGGATTTGATTTCTTTTAGGTTAGATGTAAT 834
Qy 721 AAAAAGTTAAGAAAGATACAGCTTTCCACAAAAGATTAAGAGGATATGAGTATACA 780
Db 835 AAAAAGTTAAGAAAGATACAGCTTTCCACAAAAGATTAAGAGGATATGAGTATACA 894
Qy 781 TCAAAAGTTGATTTATCAGACTCAAGCAATGTTTTCCACAAAGAGTCTTCCTCCATCA 840
Db 895 TCAAAAGTTGATTTATCAGACTCAAGCAATGTTTTCCACAAAGAGTCTTCCTCCATCA 954
Qy 841 TATGTTATTTGTTGCTACTCAAGATTTAAAGTCAAGAAATTTGGGATTTATGGAGATA 900
Db 955 TATGTTATTTGTTGCTACTCAAGATTTAAAGTCAAGAAATTTGGGATTTATGGAGATA 1014

```

Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorder; arthritis; inflammation; ss.

OS Homo sapiens.

AX  
PN  
WO200157190-A2.

XX PD 09-AUG-2001

XX  
PF 05-FEB-2001: 2001WO-US04098.XX  
PR 03-FEB-2000: 2000US-0496914-

PR 27-APR-2000; 2000US-0560875.  
PR 20-JUN-2000; 2000US-0598075

PR 19-JUL-2000; 2000US-0620325.  
PR 01-SEP-2000; 2000US-0654936

PR 15-SEP-2000; 2000US-0663561.  
PB 20-OCT-2000; 2000US-0693328

PR 30-NOV-2000; 2000US-0728422.  
yy

PA (HYSE-) HYSEQ INC.  
VV

PI Tang YT, Liu C, Drmanac RT

PI Xue A<sub>2</sub>, Yang Y, Wejhrman T

DR WPI; 2001-476283/5

XX  
XX

PT useful in diagnosis

PS Claim 1; Page 498

CC The invention relates to polynucleotides (AAK51456-AAK53435) and to

CC encoded polypeptides (AAM78323-AAM80302)  
CC cytokine, cell proliferation or cell diff

production of other cytokines in other cell populations. The polynucleotides and polypeptides are useful in gene therapy, vaccines or peptide therapy. The polypeptides have various cytokine-like activities, e.g. stem cell growth factor activity, haematopoiesis-regulating activity, tissue growth factor activity, immunomodulatory activity and activin/inhibin activity and may be useful in the diagnosis and/or treatment of cancer, leukaemia, nervous system disorders, arthritis and inflammation.

CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666  
CC (AAW80020) are omitted as the relevant pages from the sequence listing  
CC were missing at the time of publication.

Sequence 1578 BP; 496 A; 294 C; 359 G; 429 T; 0 other;

Query Match	51.1%;	Score 1464.4;	DB 22;	Length 1578;
Best Local Similarity	97.0%;	Pred. No. 0;		
Matches 1531;	Conservative	0;	Mismatches	1;
Indels	46;	Gaps	2;	

QY 1 ATGGCTCACTATATTACATTTCTCTGCAATGGTTTGGTGCTGCTTCTTCAGAAATCTGTG 60

Db  
2 ATGGCTCACTATATTACATTCTCTGCATGGTTTGGTGGCGCTTCTTCAGAAATTCCTGGC 61

61 TTAGCTGAAGATGGGGAAGTAAGATCAAGTTGTCTGGACTGCTCCGACACGATTTAGTTTC 120

1  
 2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525

100

**0**

67 GCGATATCACGAGGACCTTACACATACGAGCTTATTCAGTTCGAGTGGTTCAT 240

QY	241	TATAGTGA	CTACCTGTGCTGGAGA	TTCCTCTCGGAAGCTATGATCTACGAGAACATTTG	300
DB					
QY	242	TATAGTGA	CTACCTGTGCTGGAGAT	TCCCTCTCGGAAGCTATGATCTACGAGAACATTTG	301
DB					
QY	301	ACGCAGCA	GCTGGAATCCATACT	CTACTTAGGAGGAAACACAAGACAGGAGAGGAGGCGATC	360
DB					
QY	302	ACGCAGCA	GCTGGAATCCATACT	CTACTTAGGAGGAAACACAAGACAGGAGAGGAGGCGATC	361
DB					
QY	361	CAGTTTCG	CGCTGCAATACCTTTT	GCCAAAGTCCTCAGGATTTCTTGACTAAGATAGCAGTG	420
DB					
QY	362	CAGTTTCG	CGCTGCAATACCTTTT	GCCAAAGTCCTCAGGATTTCTTGACTAAGATAGCAGTG	421
DB					
QY	421	GTACTTAC	GCGATGGCAAAATCCC	AAAGTAGCGTCAAGGATGCGCTCAAGCAGCAAGAGAT	480
DB					
QY	422	GTACTTAC	GCGATGGCAAAATCCC	AAAGTAGCGTCAAGGATGCGCTCAAGCAGCAAGAGAT	481
DB					
QY	481	AGTAAGTA	TACANTTATTGCTATT	TGGTTGGTTTCAGAAACAGAAAGATGCCGGAACATTAGA	540
DB					
QY	482	AGTAAGTA	TACANTTATTGCTATT	TGGTTGGTTTCAGAAACAGAAAGATGCCGGAACATTAGA	541
DB					
QY	541	GCTATTGCC	CAACAAGCCTTCGCT	CTACTTATGTGCTTTTATGTGGAAGACTATATTGCAATA	600
DB					
QY	542	GCTATTGCC	CAACAAGCCTTCGCT	CTACTTATGTGCTTTTATGTGGAAGACTATATTGCAATA	601
DB					
QY	601	TCCAAA	TAAAGGGAAGTGA	TGAAGCAAACTTTGTGAAGATCTGTCTGTCCAACACGA	660
DB					
QY	602	TCCAAA	TAAAGGGAAGTGA	TGAAGCAAACTTTGTGAAGATCTGTCTGTCCAACACGA	661
DB					
QY	661	ATTCAGT	GGCAGCTGTGAT	GAAAGGGGATTTGATAATCTTTTAGTATTAGATGTAAAT	720
DB					
QY	662	ATTCAGT	GGCAGCTGTGAT	GAAAGGGGATTTGATAATCTTTTAGTATTAGATGTAAAT	721
DB					
QY	721	AAAAAG	TTAAGAAAGAAATAC	AGCTTTACCAAAAAGATATAAGATATGAAGTAACA	780
DB					
QY	722	AAAAAG	TTAAGAAAGAAATAC	AGCTTTACCAAAAAGATATAAGATATGAAGTAACA	781
DB					
QY	781	TCAAAAG	TGATTTATCAGAACT	CACAAGCAATGTTTTCCCAAGAGTCTTCCTCCATCA	840
DB					
QY	782	TCAAAAG	TGATTTATCAGAACT	CACAAGCAATGTTTTCCCAAGAGTCTTCCTCCATCA	841
DB					
QY	841	TAGTAT	TTGTGCTACTCAAGAT	TTAAAGTCAAGAAATTTGGGATTTATCGAGAAATA	900
DB					
QY	842	TATGTA	TTGTGCTACTCAAGAT	TTAAAGTCAAGAAATTTGGGATTTATCGAGAAATA	901
DB					
QY	901	TTAACT	ATTGATGGAAGGCCACA	AAATAGCAGTTTACCTTAAATCGTGTGGACAAAATCTTA	960
DB					
QY	902	TTAACT	ATTGATGGAAGGCCACA	AAATAGCAGTTTACCTTAAATCGTGTGGACAAAATCTTA	960
DB					
QY	961	TTATTT	CAACACACCGGTAAT	TATATGGCTCACAAGTGTTACCTTTGCTAACCCCTCAA	1020
DB					
QY	961	TTATTT	CAACACACCGGTAAT	TATATGGCTCACAAGTGTTACCTTTGCTAACCCCTCAA	1020
DB					
QY	1021	GTTAAG	CAGTTGTTGAT	GAAAGCTGCGACCAAAATTCGTTCTTAGTTACAGAACAGAT	1080
DB					
QY	1021	GTTAAG	CAGTTGTTGAT	GAAAGCTGCGACCAAAATTCGTTCTTAGTTACAGAACAGAT	1080
DB					
QY	1081	GTGACT	TTGTATATTGAT	GACCAACAAATTTGAAAAAGCCCTTACATCAGATTTTAGGG	1140
DB					
QY	1081	GTGACT	TTGTATATTGAT	GACCAACAAATTTGAAAAAGCCCTTACATCAGATTTTAGGG	1140
DB					
QY	1141	ATCTTGAT	CAATGGGCAACCC	CAATTTGGAAAAATATCTGGAAAAAGAAACTGTTTCAG	1200
DB					
QY	1201	TTTGGAT	GTCCAAAGTTGGGAAT	CTACTGTGACCCAGAACAGAACACCGGAGACAGCA	1260
DB					
QY	1201	TTTGGAT	GTCCAAAGTTGGGAAT	CTACTGTGACCCAGAACAGAACACCGGAGACAGCA	1260
DB					
QY	1261	TGTGAG	ATTCCTGGATTTTGC	CTTAAATGTTCCAGTGATGTAGGTTCAACTCCAGCTCCC	1320
DB					
QY	1261	TGTGAG	ATTCCTGGATTTTGC	CTTAAATGTTCCAGTGATGTAGGTTCAACTCCAGCTCCC	1320
DB					
QY	1321	TGTATT	TGTCTCGGGGAAAAC	CAGGACTTTCAAGGCCCAAGAGTGAAGCTGACCTGACATGCCT	1380
DB					

XX Nucleic acids encoding polypeptides with cytokine-like activities,  
 PT useful in diagnosis and gene therapy -  
 XX  
 PS  
 XX Claim 1; Page 4966; 6221pp; English.  
 XX  
 CC The invention relates to polynucleotides (AAK51456-AAK53435) and the  
 CC encoded polypeptides (AAK78323-AAK80302) that exhibit activity elating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activin/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation.  
 CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666  
 CC (AAK80020) are omitted as the relevant pages from the sequence listing  
 CC were missing at the time of publication.  
 XX  
 SQ Sequence 1578 BP; 496 A; 294 C; 359 G; 429 T; 0 other;

Query Match 51.1%; Score 1464.4; DB 22; Length 1578;  
 Best Local Similarity 97.0%; Pred. No. 0;  
 Matches 1531; Conservative 0; Mismatches 1; Indels 46; Gaps 2;

QY 1 ATGGCTCACTATATACATTCTCGCATGGTTTGGTCTGCTCTTCAGAAATTCGTG 60  
 DB |||||  
 QY 61 TTAGCTGAAGATGGGAAGTAAGATCAAGTTGCTGCTCCGACAGATTAGTTTC 120  
 DB |||||  
 QY 121 ATCTTAGATGCTCTTATAGTGTGGCCAGAAAATTTGAATAGTGAAGTGGCTT 180  
 DB |||||  
 QY 122 ATCTTAGATGCTCTTATAGTGTGGCCAGAAAATTTGAATAGTGAAGTGGCTT 181  
 QY 181 GTCAATATCAAAAAATTTGACATAGGCGCGAAGTTTATCAAGTTGAGTGGTCAA 240  
 DB |||||  
 QY 182 GTCAATATCAAAAAATTTGACATAGGCGCGAAGTTTATCAAGTTGAGTGGTCAA 241  
 QY 241 TATAGTGAATACCTGTGCTGGAGATTCCTCTCGGAAGTATGATTCAGAGAAATTTG 300  
 DB |||||  
 QY 242 TATAGTGAATACCTGTGCTGGAGATTCCTCTCGGAAGTATGATTCAGAGAAATTTG 301  
 QY 301 ACGGCAGCTGGAAATCCATCTACTTAGGAGGAAACACAAAGACAGGAGGCCATC 360  
 DB |||||  
 QY 361 CAGTTTGGCTCGAATACCTTTTGGCAAGTCTCTACGATTTCTGACTAAGATAGCAGT 420  
 DB |||||  
 QY 362 CAGTTTGGCTCGAATACCTTTTGGCAAGTCTCTACGATTTCTGACTAAGATAGCAGT 421  
 QY 421 GTACTTACGATGCGAAATCCCAAGATGAGCTCAAGATGCGAGCTCAAGCAGAGAGAT 480  
 DB |||||  
 QY 481 AGTAAGATAACATTTATTTGCTATTGTTGTTGTTTTCAGAAACAGAAAGATCCGAACTTAGA 540  
 DB |||||  
 QY 541 GCTATTGCCAAACAGCTTCGTCTACTTATGTTGTTTATGTTGGAAGACTATATGCAATA 600  
 DB |||||  
 QY 542 GCTATTGCCAAACAGCTTCGTCTACTTATGTTGTTTATGTTGGAAGACTATATGCAATA 601  
 QY 601 TCCAAAATAAGGGAAGTATGAAGACAGAAATTTGTGAAGAAATCTCTCTGCTCCACACGA 660  
 DB |||||  
 QY 661 ATTCAGTGGCAGCTCGTGTGAAGGGGATTTGATATCTTTTAGGTTTAGATGTAAT 720  
 DB |||||

DB 662 ATTCAGTGGCAGCTCGTGTGATGAAAGGGGATTTGATATCTTTAGGTTTAGATGTAAT 721  
 QY 721 AAAAAAGTTAAGAAAAGATAAGCTTTTCCACAAAAGATAAAAGATATAAGAGTAACA 780  
 DB |||||  
 QY 781 TCAAAAAGTTGATTTAACAAGCTCACAGCAATGTTTCCCGAGAGGCTTCTCTCCATCA 840  
 DB |||||  
 QY 841 TATGTAATTTGCTCTACTCAAGATTTAAAGTCAAGAAAATTTGGGATTTATGGAGAA 900  
 DB |||||  
 QY 901 TTAACCTATTGATGGAGGCGCACAAATAGCAGTTACTTAAATGGTGTGGAGAAAATCTTA 960  
 DB |||||  
 QY 961 TTAATTTACAAACACAGCGTAAATTAATGCTCACAAGTGGTTACCTTTGCTAACCCCTCAA 1020  
 DB |||||  
 QY 1021 GTTAAGACGTTGTTGATGAGGCTGGCACAAATTCGTCTCTTAACTACGAGCAAGAT 1080  
 DB |||||  
 QY 1081 GTGACTTTGTATATGATGACCAACAAATTTGAAAACAAAGCCCTTACATCCAGTTTGGG 1140  
 DB |||||  
 QY 1141 ATCTTATCAATGGGCAACCCAAATTTGAAAATATCTTGGAAAAGAAAGAACTCTTCAG 1200  
 QY 1201 TTTGATGTCCTCAAGATTCGCAATCTACTGTGACCCAGACAGAAACAAAGGAGACAGCA 1260  
 DB |||||  
 QY 1261 TGAGATCTCTGGATTTTGGCTTAAATGCTCCAGTGTAGTGTCAATCTCCAGCTGCC 1320  
 DB |||||  
 QY 1321 TGATTTGCTCTCCGGGAAAACACAGGACTTCAAGGCCCCAAAGGTGACCTTGGACTGGCT 1380  
 DB |||||  
 QY 1381 GGGAAACCTGGCTACCTCGCAACCTGCTCAAGATGGTAAAGCTTCAAGGCTGGCTGGCTGGCT 1440  
 DB |||||  
 QY 1441 TTAGTCACTCTCCGGTATACTGGGATACAGATATCAGGGAAATGCAAGGACACAGCT 1500  
 QY 1501 GTTCCAGGATCTCCAGGATACAGGAGCTCCAGGACTACCGGTTTACAAAGGAGAACCA 1560  
 DB |||||  
 QY 1561 GGGCGAGATGGTGACAAG 1533  
 DB |||||  
 QY 1578 GGGCGAGATGGTGACAAG 1578  
 DB |||||  
 RESULT 4  
 ID AAK53249  
 XX AAK53249 standard; cDNA; 1578 BP.  
 AC AAK53249;  
 XX AC  
 DT 06-NOV-2001 (first entry)  
 XX  
 DE Human polynucleotide SEQ ID NO 2778.  
 XX

181 GTCAATATCAAAAAAATTTTGACATAGGGCCGAAAGTTTATTCAAGTTGGAGTGGTTCAA 240  
 Db |||||  
 182 GTCAATATCAAAAAAATTTTGACATAGGGCCGAAAGTTTATTCAAGTTGGAGTGGTTCAA 241  
 Qy |||||  
 241 TATAGTACTACCTGTGTGGAGATTCTCTCGGAAGCTATGATTCAGGAGAACATTGG 300  
 Db |||||  
 242 TATAGTACTACCTGTGTGGAGATTCTCTCGGAAGCTATGATTCAGGAGAACATTGG 301  
 Qy |||||  
 301 ACGGACGAGTGGATTCATCTACTTACGAGGAAACACAAAGACAGGAGGAGCCATC 360  
 Db |||||  
 302 ACGGACGAGTGGATTCATCTACTTACGAGGAAACACAAAGACAGGAGGAGCCATC 361  
 Qy |||||  
 361 CAGTTTGGCTGCTGATTAACCTTTTGGCAAGTCTCTACGATTTCTGACTAAGATAGCAGTG 420  
 Db |||||  
 362 CAGTTTGGCTGCTGATTAACCTTTTGGCAAGTCTCTACGATTTCTGACTAAGATAGCAGTG 421  
 Qy |||||  
 421 GTACTTACGGATGGCAAAATCCCAAGATGACGTCAAGGATGCGAGTCAAGCAGCAAGAGAT 480  
 Db |||||  
 422 GTACTTACGGATGGCAAAATCCCAAGATGACGTCAAGGATGCGAGTCAAGCAGCAAGAGAT 481  
 Qy |||||  
 481 AGTAAGATACATTTATTTGCTATTGGTGTGGTTCAGAAACAGAGATGCCAACTTAGA 540  
 Db |||||  
 482 AGTAAGATACATTTATTTGCTATTGGTGTGGTTCAGAAACAGAGATGCCAACTTAGA 541  
 Qy |||||  
 541 GCTATTGCCAACAGGCTTCGCTACTTATGTGTTTATGTGGAAGACTATATTGCAATA 600  
 Db |||||  
 542 GCTATTGCCAACAGGCTTCGCTACTTATGTGTTTATGTGGAAGACTATATTGCAATA 601  
 Qy |||||  
 601 TCCAAATTAAGGAAGTGAATGAAGCAAACTTTGTGAAGATCTGTCTCCACACGA 660  
 Db |||||  
 602 TCCAAATTAAGGAAGTGAATGAAGCAAACTTTGTGAAGATCTGTCTCCACACGA 661  
 Qy |||||  
 661 ATTCCAGTGGCAGCTCGTGATGAAGCGGATTTGATTTCTTTTAGTTTACATGTAAT 720  
 Db |||||  
 662 ATTCCAGTGGCAGCTCGTGATGAAGCGGATTTGATTTCTTTTAGTTTACATGTAAT 721  
 Qy |||||  
 721 AAAAAAGTTAAGAAAAAGATACAGCTTTCACCAAAAAAGATAAAGGATATGAAGTAACA 780  
 Db |||||  
 722 AAAAAAGTTAAGAAAAAGATACAGCTTTCACCAAAAAAGATAAAGGATATGAAGTAACA 781  
 Qy |||||  
 781 TCAAAAGTTGATTTATCAGAACTCACAAGCAATGTTTCCGAGAGGCTTCTCTCATCA 940  
 Db |||||  
 782 TCAAAAGTTGATTTATCAGAACTCACAAGCAATGTTTCCGAGAGGCTTCTCTCATCA 941  
 Qy |||||  
 841 TATGATTTGTGCTACTCAAGATTTAAAGTCAAGAAATTTGGATTTATGGAGATA 900  
 Db |||||  
 842 TATGATTTGTGCTACTCAAGATTTAAAGTCAAGAAATTTGGATTTATGGAGATA 901  
 Qy |||||  
 901 TTAACTATTGATGGAAGGCCACAAATAGCAGTTTACCTTAAATGGTGGACAAAATCTTA 960  
 Db |||||  
 902 TTAACTATTGATGGAAGGCCACAAATAGCAGTTTACCTTAAATGGTGGACAAAATCTTA 961  
 Qy |||||  
 961 TTATTTACAAACACAGCGTAATTTAATGGCTCAAGTGGTTACCTTTGCTAACCTCAA 1020  
 Db |||||  
 962 TTATTTACAAACACAGCGTAATTTAATGGCTCAAGTGGTTACCTTTGCTAACCTCAA 1021  
 Qy |||||  
 1021 GTTAAGACGTTGTTGATGAAGGCTGGCCACCAATTCGTTCTTTAGTAACAGAACAGAT 1080  
 Db |||||  
 1022 GTTAAGACGTTGTTGATGAAGGCTGGCCACCAATTCGTTCTTTAGTAACAGAACAGAT 1081  
 Qy |||||  
 1081 GTGACCTTGTATTTATGATGACCAACAAATTTGAACAAAGCCCTTACATCCAGTTTAGGG 1140  
 Db |||||  
 1082 GTGACCTTGTATTTATGATGACCAACAAATTTGAACAAAGCCCTTACATCCAGTTTAGGG 1141  
 Qy |||||  
 1141 ATCTTGATCAATGGGCAACCCCAATTTGGAATAATTTCTGAAAAAGAAACTGTTAG 1200  
 Db |||||  
 1142 ATCTTGATCAATGGGCAACCCCAATTTGGAATAATTTCTGAAAAAGAAACTGTTAG 1201  
 Qy |||||  
 1201 TTTGATGTCACCAAAAGTTGGAATCTACTGTGACCCAGAAACAGAACACCGGGAGACGA 1260  
 Db |||||  
 1202 TTTGATGTCACCAAAAGTTGGAATCTACTGTGACCCAGAAACAGAACACCGGGAGACGA 1261

1261 TGTGAGATTCTCGATTTGCTTTAATGTTCCAGTGATGATAGGTTCAACTCCAGTCCC 1320  
 Db |||||  
 1262 TGTGAGATTCTCGATTTGCTTTAATGTTCCAGTGATGATAGGTTCAACTCCAGTCCC 1321  
 Qy |||||  
 1321 TGTATTTGCTCTCGGGAACCAAGGACTTCAAGGCCCCAAAGGTGACCTTGACTGGCT 1380  
 Db |||||  
 1322 TGTATTTGCTCTCGGGAACCAAGGACTTCAAGGCCCCAAAGGTGACCTTGACTGGCT 1381  
 Qy |||||  
 1381 GGGAAACCTGCTACCTGGACAACTGCTCAAGATGGTAAGCT----- 1425  
 Db |||||  
 1382 GGGAAACCTGCTACCTGGACAACTGCTCAAGATGGTAAGCTCAAGTACTGAAAGC 1441  
 Qy |||||  
 1426 -----GGATACAGGAAATTCAGGACACACAGGT 1455  
 Db |||||  
 1442 TTAGTCATCTCCGTTATCTCGGATACAGGATATCAGGAAATTCAGGACACACAGGT 1501  
 Qy |||||  
 1456 GTTCCAGGATCTCAGAAATACAGGAGCTCGAGGACTACAGGTTACAAAGGAGAACCA 1515  
 Db |||||  
 1502 GTTCCAGGATCTCAGAAATACAGGAGCTCGAGGACTACAGGTTACAAAGGAGAACCA 1561  
 Qy |||||  
 1516 GGGCGAGATGTCACAAAGGCTGATCGTGAGCTTCTGTTTCTCGGCTTCATGGCATG 1575  
 Db |||||  
 1562 GGGCGAGATGTCACAAAGGCTGATCGTGAGCTTCTGTTTCTCGGCTTCATGGCATG 1621  
 Qy |||||  
 1576 CCAGGATCAAGAGGTGAATGGTGCCAAAGSAGACAAAGGATCACTGGATTTATGGC 1635  
 Db |||||  
 1622 CCAGGATCAAGAGGTGAATGGTGCCAAAGSAGACAAAGGATCACTGGATTTATGGC 1681  
 Qy |||||  
 1616 AAAAAAGGTCACAAAGGTGAAA 1657  
 Db |||||  
 1682 AAAAAAGGAGATAGATCACA 1703

RESULT 3  
 AAK53227 0  
 ID AAK53227 standard; cDNA: 1578 BP.  
 AC AAK53227;  
 XX  
 DT 06-NOV-2001 (first entry)  
 XX  
 DE Human polynucleotide SEQ ID NO 2756.  
 XX  
 DE Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 KW vaccine; peptide therapy; stem cell; growth factor; haematopoiesis;  
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
 KW nervous system disorder; arthritis; inflammation; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WC200157190.A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 05-FEB-2001; 2001WO-US04058.  
 XX  
 PR 03-FEB-2000; 2000US-0496914.  
 PR 27-APR-2000; 2000US-0560875.  
 PR 20-JUN-2000; 2000US-0598075.  
 PR 19-JUL-2000; 2000US-0620325.  
 PR 01-SEP-2000; 2000US-0654936.  
 PR 15-SEP-2000; 2000US-0663561.  
 PR 20-OCT-2000; 2000US-0693325.  
 PR 30-NOV-2000; 2000US-0728422.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;  
 PI Zhao QA, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
 PI Xue AJ, Yang Y, Wehrman T, Goodrich R;  
 XX  
 DR WPI; 2001-476283/51.  
 DR P-PSDB: AAK90094.



PT different human tissues, useful in recombinant DNA methodologies -  
 XX Claim 1; Page 201-202; 1095pp; English.  
 XX This invention describes novel polynucleotides and polypeptides isolated from human cDNA libraries which can be used for gene therapy or in vaccines. The polynucleotides of the invention and antibodies encoded by them may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate polypeptide expression. The products of the invention may also be used to identify modulators of expression and activity and to down regulate expression and activity. The antibodies of the invention may also be used as diagnostic agents for detecting the presence of polypeptides in samples. This sequence encodes a polypeptide described in the disclosure of the invention.

SQ Sequence 4160 BP; 1271 A; 795 C; 1026 G; 1068 T; 0 other;

Query Match 99.1%; Score 2839.6; DB 23; Length 4160;  
 Best Local Similarity 99.5%; Pred. No. 0;  
 Matches 2861; Conservative 0; Mismatches 4; Indels 9; Gaps 1;

QY 1 ATGGCTCACTATATACATTTCTCTGCGATGTTTGGTCTCTTTCAGAAATCTGTG 60  
 DB 203 ATGGCTCACTATATACATTTCTCTGCGATGTTTGGTCTCTTTCAGAAATCTGTG 262  
 QY 61 TTAGCTGAAGATGGGGAAGTAAAGTCAAGTTCTGCTACTGCTCCGACAGATTTAGTTTC 120  
 DB 263 TTAGCTGAAGATGGGGAAGTAAAGTCAAGTTCTGCTACTGCTCCGACAGATTTAGTTTC 322  
 QY 121 ATCTTAGATGGCTCTTATAGTGTGGCCAGAAAACCTTTGAAATAGTGAAGAGTGGCTT 180  
 DB 323 ATCTTAGATGGCTCTTATAGTGTGGCCAGAAAACCTTTGAAATAGTGAAGAGTGGCTT 382  
 QY 181 GTCATATCACAAAACCTTTCAGATAGGCGGAGATTTATTCAGTTGGAGTGGTCAA 240  
 DB 383 GTCATATCACAAAACCTTTCAGATAGGCGGAGATTTATTCAGTTGGAGTGGTCAA 442  
 QY 241 TATAGTACTACCTGTGCTGAGATTTCTCTCGAAGCTATGATTCAGGAGAACATTG 300  
 DB 443 TATAGTACTACCTGTGCTGAGATTTCTCTCGAAGCTATGATTCAGGAGAACATTG 502  
 QY 301 AGGCGAGAGTGGATTCATCTACTCTAGGAGGAAACACAAAGAGAGGAGGCCATC 360  
 DB 503 AGGCGAGAGTGGATTCATCTACTCTAGGAGGAAACACAAAGAGAGGAGGCCATC 562  
 QY 361 CAGTTTGGCTCGATTTACCTTTTGGCAAGTCTTCCAGATTTCTCACTAAGATAGCAGTG 420  
 DB 563 CAGTTTGGCTCGATTTACCTTTTGGCAAGTCTTCCAGATTTCTCACTAAGATAGCAGTG 622  
 QY 421 GTACTTACGGATGGCAATCCCAAGATGACGTCAGAGTGGCAAGTGGCAAGTGGCAAT 480  
 DB 623 GTACTTACGGATGGCAATCCCAAGATGACGTCAGAGTGGCAAGTGGCAAGTGGCAAT 682  
 QY 481 AGTAAGATAACATTTTCTATTTGTTTATGTTTATGTTGGAAGACTATATTCGAATA 540  
 DB 683 AGTAAGATAACATTTTCTATTTGTTTATGTTTATGTTGGAAGACTATATTCGAATA 742  
 QY 541 GCTATTGCCAAACAAGCTTCTGCTACTTATGTTTATGTTGGAAGACTATATTCGAATA 600  
 DB 743 GCTATTGCCAAACAAGCTTCTGCTACTTATGTTTATGTTGGAAGACTATATTCGAATA 802  
 QY 601 TCCAAATTAAGGGAAGTATGAGGAGAACTTTGTGAAGATCTGTCGTCCACAGCA 660  
 DB 803 TCCAAATTAAGGGAAGTATGAGGAGAACTTTGTGAAGATCTGTCGTCCACAGCA 862  
 QY 661 ATTCCAGTGGAGCTGCTGATGAAAGGGATTTGATATTTCTTTAGTTTGAATTAAT 720  
 DB 863 ATTCCAGTGGAGCTGCTGATGAAAGGGATTTGATATTTCTTTTGGGTTTGAATTAAT 922  
 QY 721 AAAAAGGTTAAGAAAGAAATACAGCTTTTCCAAAAAAGATAAAGAGGATTAAGTAA 780  
 DB 923 AAAAAGGTTAAGAAAGAAATACAGCTTTTCCAAAAAAGATAAAGAGGATTAAGTAA 982

QY 781 TCAAAAGTTGATTTATCAGAACTCACAAGCAATGTTTTCAGAGAGTCTTCTCCATCA 840  
 DB 983 TCAAAAGTTGATTTATCAGAACTCACAAGCAATGTTTTCAGAGAGTCTTCTCCATCA 1042  
 QY 841 TATGATTTGTGTCTACTCAAGATTTAAAGTCAAGAAAATTTGGGATTTATGGAGAATA 900  
 DB 1043 TATGATTTGTGTCTACTCAAGATTTAAAGTCAAGAAAATTTGGGATTTATGGAGAATA 1102  
 QY 901 TTAACATTTGATGGRAGGCCACAAATAGCAGTTACCTTAAATGTTGTGTGAGCAAACTTA 960  
 DB 1:03 TTAACATTTGATGGAAGGCCACAAATAGCAGTTACCTTAAATGTTGTGAGCAAACTTA 1162  
 QY 961 TTATTTACAAACACCAGCGTAATTAATGCTCACAAGTGTGTTACCTTTTAACTTAACT 1020  
 DB 1:63 TTATTTACAAACACCAGCGTAATTAATGCTCACAAGTGTGTTACCTTTTAACTTAACT 1222  
 QY 1021 GTTAAGAGCTGTTTGTATGAGGCTGGCACCATAATTCGTCTCTTAGTAAACAGAAAT 1050  
 DB 1223 GTTAAGAGCTGTTTGTATGAGGCTGGCACCATAATTCGTCTCTTAGTAAACAGAAAT 1282  
 QY 1081 GTGACTTTGTATCTGATGACCAACAAATTTGAAACCAAGCCCTTACATCCAGTTTATGG 1140  
 DB 1:83 GTGACTTTGTATCTGATGACCAACAAATTTGAAACCAAGCCCTTACATCCAGTTTATGG 1342  
 QY 1:41 ATCTTGATCAATGGCCAAACCCAAATTTGAAATAATTTCTGAAAGAAAGAACTGTTCAG 1200  
 DB 1343 ATCTTGATCAATGGCCAAACCCAAATTTGAAATAATTTCTGAAAGAAAGAACTGTTCAG 1402  
 QY 1201 TTTGATGTCACAAAGTTGGAATCTACTGTGACCCAGAAACAGNACACCGGAGAGACGA 1260  
 DB 1403 TTTGATGTCACAAAGTTGGAATCTACTGTGACCCAGAAACAGNACACCGGAGAGACGA 1462  
 QY 1:261 TGTGAGATTCCTGSAATTT-----TGCTTAATGCTGCTGCTGCTGCTGCTGCTGCT 1311  
 DB 1463 TGTGAGATTCCTGSAATTTAATGAGAGTGGCTTAAATGCTGCTGCTGCTGCTGCTGCT 1522  
 QY 1312 CAGCTCCCTGTATTTGTCTCCGGGAAACCCAGGACTTCAAGGCCCCCAAGAGTGAACCT 1371  
 DB 1:523 CAGCTCCCTGTATTTGTCTCCGGGAAACCCAGGACTTCAAGGCCCCCAAGAGTGAACCT 1582  
 QY 1372 GGAAGTCCGGGAAACCCCTGCTACCTGTCGACCACTGTCGAGATGAGTGAAGCTGATAT 1431  
 DB 1583 GGAAGTCCGGGAAACCCCTGCTACCTGTCGACCACTGTCGAGATGAGTGAAGCTGATAT 1642  
 QY 1432 CAGGAAATTCAGGGACACCAAGTGTTCAGAGTCTCCAGGATAAAGAGAGTCCAGGA 1491  
 DB 1643 CAGGAAATTCAGGGACACCAAGTGTTCAGAGTCTCCAGGATAAAGAGAGTCCAGGA 1702  
 QY 1:492 CTACAGGTTTCAAGAGAGAACCAAGGCGAGATGTCGACCAAGGCTGATGCTGGACTTCT 1551  
 DB 1703 CTACAGGTTTCAAGAGAGAACCAAGGCGAGATGTCGACCAAGGCTGATGCTGGACTTCT 1762  
 QY 1552 GGTCTTCTCCGGCTTATGCTGTCAGGATCAAGGGTGAATGCTGCTGCTGCTGCTGCTGCT 1611  
 DB 1763 GGTCTTCTCCGGCTTATGCTGTCAGGATCAAGGGTGAATGCTGCTGCTGCTGCTGCT 1822  
 QY 1612 AAAGGATCACCTGGATTTTATGGCAAAAAGGGTGAATGCTGCTGCTGCTGCTGCTGCT 1671  
 DB 1823 AAAGGATCACCTGGATTTTATGGCAAAAAGGGTGAATGCTGCTGCTGCTGCTGCTGCT 1982  
 QY 1672 TTCCCTGGCTTCCCTGCACTGCTGAGAACCAAGGAGACATGGAAGAGATGATTAATG 1731  
 DB 1883 TTCCCTGGCTTCCCTGCACTGCTGAGAACCAAGGAGACATGGAAGAGATGATTAATG 1942  
 QY 1732 GGTAGTCCCGGTTTCAAGGAGAGAACGAGGATCCCTGCTGCTGCTGCTGCTGCTGCTGCT 1791  
 DB 1943 GGTAGTCCCGGTTTCAAGGAGAGAACGAGGATCCCTGCTGCTGCTGCTGCTGCTGCT 2002  
 QY 1792 CGGGGAGAGCTGGAATCCAGGATTTCTCTGGAACCCAGGATTAATGCTGCTGCTGCTGCT 1851  
 DB 2003 CGGGGAGAGCTGGAATCCAGGATTTCTCTGGAACCCAGGATTAATGCTGCTGCTGCTGCT 2062  
 QY 1852 GAAATTTGGGCTTCCAGGACAGCAAGGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1911

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2003, 18:17:32 ; Search time 715 Seconds  
(without alignments)

10816.629 Million cell updates/sec

Title: US-09-996-611B-5

Perfect score: 2865

Sequence: 1 atgctcactatattacatt.....gaaagacacaaactattag 2865

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 252756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq : 3Jun03.\*  
1: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*  
25: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2839.6	99.1	4160	23	ABX71234
2	1589.2	55.5	1826	22	AAK52243
3	1464.4	51.1	1578	22	AAK53227
4	1464.4	51.1	1578	22	AAK53249
5	1462.8	51.1	1691	22	AAK52265
6	1282	44.7	2375	22	AAK526053
7	1282	44.7	2375	25	ABX73394
8	1062	37.1	2230	22	AAK52045

9	1062	37.1	2230	22	AAI58402
10	1039.8	36.3	2209	22	AAK53029
11	1039.8	36.3	2209	22	AAI60188
12	974	34.0	1184	24	ABQ72640
13	972.4	33.9	1169	24	ABQ72511
14	908	31.7	1115	22	AAK31132
15	453.6	15.8	1414	24	ABK71703
16	353.4	12.3	447	20	AAV86274
17	285	9.9	6292	25	ACC00270
18	255	9.9	6352	25	ACC00269
19	268.2	9.4	4428	22	AD06578
20	259.8	9.1	4428	22	AD06574
21	259.8	9.1	4428	22	AD06575
22	242.6	8.5	546C	17	AAI16508
23	242.6	8.5	546C	22	ABA83117
24	242.6	8.5	5460	24	ABL92101
25	242.6	8.5	5460	25	ACC50110
26	242.6	8.5	5460	25	ABX72026
27	242.6	8.5	5466	23	AAK79378
28	242.6	8.5	5489	24	ABX04154
29	242.4	8.5	5467	22	AAH98343
30	242.4	8.5	5468	22	AAH98411
31	241.4	8.4	6319	24	ABX04156
32	239.8	8.4	6158	24	ABT10872
33	239.8	8.4	6158	24	ABT07753
34	239.8	8.4	6158	24	ABL62095
35	239.8	8.4	6158	24	ABL65452
36	239.9	8.4	6158	24	ABK35486
37	239.8	8.4	6158	25	ABX76310
38	236.6	8.3	4816	24	ABK84041
39	236.6	8.3	6355	24	ABQ60877
40	236.6	8.3	6691	23	AAK79806
41	235.8	8.2	6882	20	AAZ10631
42	235.2	8.2	4172	22	AAH62769
43	235.2	8.2	305107	22	AAH62689
44	234.4	8.2	3690	13	AAQ30849
45	233.8	8.2	5058	23	AAK69026

#### ALIGNMENTS

#### RESULT 1

ABX71234  
ID ABX71234 standard; cDNA; 4160 BP.

XX AC ABX71234;

XX DT 15-APR-2003 (first entry)

XX DE Cell structure and mobility-associated cDNA from clone DKFZphfbr2\_2b5.

XX KW Human; gene; gene therapy; vaccine; disease treatment; detection; ss.

XX OS Homo sapiens.

XX PN NC200112659-A2.

XX PC 22-FEB-2001.

XX PF 18-AUG-2000; 2000MO-2B01496.

XX PR 18-AUG-1999; 99US-0149499.

XX PR 28-SEP-1999; 99US-0156503.

XX (GEHU-) GERMAN HUMAN GENOME PROJECT.

XX WIemann S;

XX DR WP1; 2001-327840/34.

XX DR P-PSDB; ABU52683.

XX PT Nucleic acids having the sequences of clones isolated from libraries of

Db	1875	TGAACCCGGCTTACCTGGACCCACAGGTTACATTGGGCCACAGGGTTCGTGTTGGTCTCC	1934
Qy	2790	AGGCATCTGGGACCCGATCACTATGTTTTAGTGTAAATTGCCGAGAAGAGATCCGTTCAAGAAA	2849
Db	1935	TGGTTACTGGGACCCGATCACTTGTGTTGATGGGTCAAGTCCGTTGGCATTTCAACAACCTAA	1994
Qy	2850	AGGACCAAA	2858
Db	1995	AGGTCCAAA	2003

Search completed: October 2, 2003, 22:19:10  
Job time : 10:51 secs



```

Db      2732 CTATGGGAAATGCGGAGACGGTGAGGAGGCCCCCAGGGCTGGCAGGAATCTCTGG 2791
QY      2757 AAAAAGTGAATCAAGGCAACAGGCCCCCAGGCATCTCGACCCATC 2807
Db      2792 AGTGCTGGACCCCGGACCTCTGGGCTTCCCGTTCTGTGAGCCAGC 2842

RESULT 15
AK115182
LOCUS   AK115182                2359 bp    mRNA    linear    INV 30-NOV-2002
DEFINITION   Ciona intestinalis cDNA, clone:cilv0.4e05, full insert sequence.
ACCESSION   AK115182
VERSION     AK115182.1    GI:23587129
KEYWORDS    FLI CDNA.
SOURCE      Ciona intestinalis
ORGANISM    Ciona intestinalis
Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogoga;
Phlebobranchia; Clonidae; Ciona.

REFERENCE
1
AUTHORS    Satou,Y., Yamada,L., Mochizuki,Y., Takatori,N., Kawashima,T.,
            Sasaki,A., Hanaguchi,M., Awazu,S., Yagi,K., Sasakura,Y.,
            Nakayama,A., Ishikawa,H., Inaba,K. and Satoh,N.
TITLE      A cDNA resource from the basal chordate Ciona intestinalis
JOURNAL    Genesis 33 (4), 153-154 (2002)
MEDLINE    22191024
PUBMED     12203911
REFERENCE  2 (bases 1 to 2359)
AUTHORS    Satou,Y. and Satoh,N.
DIRECT SUBMISSION
SUBMITTED  (04-OCT-2002) Nori Satoh, Kyoto University, Department of
Zoology, Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
            (E-mail:satoh@sci.kyoto-u.ac.jp, Tel:81-75-753-4095,
            Fax:81-75-705-1113)
COMMENT    Ciona intestinalis cDNA Project (URL:
            http://ghost.zool.kyoto-u.ac.jp/indexr1.html).
FEATURES
source     1..2359
            /organism="Ciona intestinalis"
            /mol_type="mRNA"
            /db_xref="taxon:7719"
            /clone="cilv0.4e05"
BASE COUNT 641 a 552 c 724 g 442 t

Query Match      10.2%; Score 292.6; DB 3; Length 2359;
Best Local Similarity 50.3%; Pred. No. 3.5e-55;
Matches 789; Conservative 0; Mismatches 759; Indels 21; Gaps 2;

QY      1299 TGATGTTCAATCCAGCTCCCTGTATTTGTCTCCGGGAAACCCAGGACTTCAAGGCC 1358
Db      447 TGGAGGCGTACGGCTGGTGAAGAGTGGTTTCAAAATGGTGAAGTTGGTTTGTGGACC 506

QY      1359 CAAAGGTGACCTGGAGCTCCCTGGGAAACCTCGCTACCTCGACACCTGGTCAAGATGG 1418
Db      507 ACCAGGTTATCCCGACCCAGGTCATAAGGATACAAAGTGAAACCCGGTCTCTGATGG 566

QY      1419 TAAGCTTGATATCAGGGAATTCAGGGACACAGGTGTTCAGGATCTCCAGGAATACA 1478
Db      567 AAAACAAGTCCAAAGGGTGATCCCGTCTCGAGGAGAACCAAGCAGCTGGTCTTGG 626

QY      1479 AGGAGCTCAGGAGTACAGGTTACAAAGAGAACCCAGGCGAGATGGTGAACAAGGTGA 1538
Db      627 CGGTGAATCTGTGTCCACGTGGGACAGAGGTTACCCGGGAGAGATGGGAGAAAGGAGA 686

QY      1539 TCGTGACTTCTCTGGTTTCTCTGGGCTTCATGSCATGCCAGGATCAAGGGTGAATGGG 1599
Db      687 AGTAGGCCCTGCAGGAGTACCTGGAGAAATGGGTTCACTTGTGTGAACAGGTTTACAGGG 746

QY      1599 TGCCAAAGGAGACAAAGGATCACCTGGATTTTATGGCAAAAAGGGTGCAAAAGGTGA 1659
Db      747 AGCTGAAGTCTCTCTGGAGTCCAGGAGAGGGGAGCCAGAGGAGGAAGTGGGGATCG 806

1659 GGGAAATGCTGGCTTCCCTGGCTCCTGGACCTGCTGGAGAACCCAGGAAGACATGAAA 1718
807 AGGTCTCTTTGGAAAACCTGGAGAACCGGACTTCTGGAGAACCCGGTCTACCCGGTAG 866
1719 GGATGAAATTAATGGTAGTCCCGTTTCAAGGGAGAACGAGGATCCCTCTGGTCTCCGG 1778
867 AATGGAGCTGATGTTATCCAGGAACCCAGGAAACCGGGAGACCCAGGAGAGTCAAG 926
1779 GAGGATGGAAACAGGGGAGAGCTGGAATCCAGGATTTCTTGGAAACCGAGGATTAAT 1838
927 ACCAAAGGAACAGAGGTCCTATAGAGCTCTTGGAAATACCTGGAGATATTTGAGCTGA 986
1839 GGGCAAAAGGAGAAATTTGGGCTCCAGSACAGCAAGGAAAAAAGAGAGCC----- 1891
987 AGGACCAACTGGCAACAGCTGGTGCACAGGTAATCCAGGAGAGAATGGCGCCGTGACC 1046
1892 --CAGGATCCCTGGTTTAAATGGGAAGCAATGGCTCACAGGCCAGCTGGAAACACGGG 1949
1047 AAAAGGAGAAAGTGGGACCAATTTGGACAAATTTGAGTTACAGGTCTCAGCTGGAGAGG 1106
1950 ATCTAAGGGAAGCAAGGTGAACCTGGAATTTCAAGGATGCCCTGGGGCTTCTGGGCTCAA 2009
1107 ACAGTGGAGATGAGGACCATCAGAGAAATCTGACCACTTGGACCCCGTGGTAACAA 1166
2310 GGGAGAACCAAGGACCAACGGGTTCCCAAGAGAACCCAGGATACATGGGTTTACCCGGAT 2069
1167 GGGTAATTCAGGAGAACAGGTATTTCAGGCTCAACCTGGTGAACGTGGATTCCTCTGAA 1226
2670 TCAAGGAAAGGGGAGCAAGGAAATCAAGTGAAAGGCTATTCAGGTCAAAGGG 2129
12276 CAAAGGACACAAAGGAGCATCTGTATGACTGGCGCCAAAGGAGAACCTGGTAAATGATGG 1286
2130 AGAAATGGAGACAGGGAATTCAGGCGAAACAGGAATTTCAAGGCCATTCATGGTCAAA 2189
1287 AATGAGGACACAGAGGCAACCTCGACCAAGGAACTGTGGAACTCCAGGAGAACG 1346
2190 AGGAGAGAGAGTGAAAGGGAGAACCTGGTGTCCGAGGTGCGCATTTGGATCAAAAGGAG 2249
1347 TGGATCACAGGAACACCGGGTGAAATGGGCAACCTGGCCAGGTGGACCCAGCGGCA 1406
2250 ATCTGGGTGGATGGCTTGTATGGGCGCCGAGCTCTTAAGGGCAACCTGGGGATCCAGG 2309
1407 CATTTGACCAAGAGGACCCAGGGGCCAATCCGGGCCACAGGAGCACCAAGGAGAACTGG 1466
2310 TCTCTAGGACCCCGCAGGTTTGGATGGAGACCCGGAAGAGATTTTTCAGAACAATTTAT 2369
1467 GCCAAGAGGCGTCCCTGGATTACAAGCCCTCTCTGGCGCATGCCAAGTGACCAAGAT 1526
2370 TCGACAAGTTTGCACAGATGTAATAAGAGCCAGCTACCAAGTCTTACTTTCAGAGTGGAG 2429
1527 TCTGAAATTTGGCGTAGGGTGTTCGAGATGAGATGAGAGGTATGATGGCAGCAACCCAG 1586
2430 AATTAGAAATTTGTGATCATTTGCTGCTGCCAACATGCTCCCGGCTATTTCTTGGGCCACC 2489
1587 ACACAGGATCATTTACTG-----TGGAGCTCTCTGGAAGACCGGACCTGT 1634
2490 TGTTCGATASGCGCCAGAGGTTCCAGAGATTTACCTGGTTTCCAGGAGAGAGATGGTGT 2549
1635 TGGACCACTTGGACAGCAAGGACCAACAGCAATCTGTCTTCTGTGTAAATCGAAT 1694
2550 TCCTGGATTTAGTGGTGTCTCTGGAGCTCCAGGTGTCAGAGGTTTAAAGGCGCTACAGG 2609
1695 GCAAGGTGCAACGGGTGTTCTTGGAAATGCTTGGACGTACTGACCCCAAGGATTTGGTGG 1754
2610 AAGAAATGGGAAAAAGGAGGCGCAAGGTTTGGGTATCTCTGGAGAACAGGTCTCTCTGG 2669
1755 TGATAAGAGTGAAGAGAGGCGCGGAAATTTGACAGCTTGGTCTTCCAGGACCTGCTGG 1814
2670 TCCCGGAGTCCAGAGGCGCTCTCTGGAATTAAGCAAGAGGTCCTCCAGGAGACCCAGG 2729
1815 TCTTCAGGACCCCAAGGATCACCGGTTTGGCAAGATGGTCTGTAAACGGAGACCTGG 1874
2730 TCTCCTCTGGCAAGATGGAGACCATGGAAATCTGGAAATTCAGGAGCAACCAAGGCCCCC 2789

```



DB 1223 AGGTCCTCCTGGAGTGCCTGGGTTTGATGGGAAATCTGGCAAAACCGGTGAACAGGGACC 1282

```

source
1. 1414
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/ncbi="Incyte ID No: LG:407788.2:2000SEP08"
BASE COUNT 425 a 267 c 295 g 427 t
ORIGIN

Query Match 15.8%; Score 453.6; DB 6; Length 1414;
Best Local Similarity 99.1%; Pred. No. 1.2e-91; Indels 0; Gaps 0;
Matches 456; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2393 TAAGAGCCAGCAGCAGCTTCTTACAGAGTGAAGAAATAGAAATTTGTGATCATGGCC 2452
DB 955 TGAAGCCAGCAGCAGCTTCTTACAGAGTGAAGAAATAGAAATTTGTGATCATGGCC 10.4
QY 2453 TGTCCCAATGCTCCCGGGTATCTTGGCCACCTGGTCCGATAGGCCAGAGGGTC 2512
DB 1015 TGTCCCAATGCTCCCGGGTATCTTGGCCACCTGGTCCGATAGGCCAGAGGGTC 1074
QY 2513 CCAGAGGATTACCTGGTTTCCAGGAAGAGATGGTTCCTCGGATTAGTGGTGTCCCTG 2572
DB 1075 CCAGAGGATTACCTGGTTTCCAGGAAGAGATGGTTCCTCGGATTAGTGGTGTCCCTG 1134
QY 2573 GACGTCCAGGTGTGAGAGATTAAAGGCTTACAGGAAGAAATGGGAAAGAGAGCC 2632
DB 1135 GACGTCCAGGTGTGAGAGATTAAAGGCTTACAGGAAGAAATGGGAAAGAGAGCC 1194
QY 2633 AAGGTTTGGGTATCTCCGAGAACCAAGTCTCTGTCCTCCAGGTCACAGAGGCCCTC 2692
DB 1195 AAGGTTTGGGTATCTCCGAGAACCAAGTCTCTGTCCTCCAGGTCACAGAGGCCCTC 1254
QY 2693 CTGGAATAAGCAAGAAGTCTCCAGGAGACCCAGGTCTCTCCCTGGCAAGATGGAGCC 2752
DB 1255 CTGGAATAAGCAAGAAGTCTCCAGGAGACCCAGGTCTCTCCCTGGCAAGATGGAGCC 1314
QY 2753 ATGGAACCTGGATCCAGGCCAAGCCAGAGCCGCCAGGATCTGGACCCATCATAT 2812
DB 1315 ATGGAACCTGGATCCAGGCCAAGCCAGAGCCGCCAGGATCTGGACCCATCATAT 1374
QY 2813 GTTTTAGTGTAAATGCCAGAGAGATCCGTTCCAGAAAAGG 2852
DB 1375 GTTTTAGTGTAAATGCCAGAGAGATCCGTTCCAGAAAAGG 1414

RESULT 12
BD058397
LOCUS
DEFINITION
SECRETED expressed sequence tags (ESTs).
ACCESSION
BD058397
VERSION
BD058397.1 GI:22604003
KEYWORDS
JP 2001519666-A/252.
SOURCE
Zea mays
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE
1. (bases 1 to 447)
JACOBS,K., MCCOY,J.M., LAVALLIE,E.R., RACIE,L.A., MERBERG,D.,
TREACY,M., SPAULDING,V. and AGOSTINO,M.J.
Secreted expressed sequence tags (ESTs)
Patent: JP 2001519666-A 252 23-OCT-2001;
GENETICS INSTITUTE INC
COMMENT
PN JP 2001519666-A/252
PD 23-OCT-2001
PF 10-APR-1998 JP 1998543068
PR 10-APR-1997 US 08/835913
PI KENNETH JACOBS,JOHN M MCCOY,EDWARD R LAVALLIE,LISA A RACIE, PI
DAVID MERBERG,
PI MAURICE TREACY,VIKKI SPAULDING,MICHAEL J AGOSTINO PC
C12N15/12,C12N5/10,C07K14/47,C12Q1/68,A61K38/17 CC Strandedness:
Double;
CC Topology: Linear;

FEATURES
source
1. 447
/organism="Zea mays"
/mol_type="genomic DNA"
/db_xref="taxon:4577"
BASE COUNT 141 a 141 c 101 g 125 t
ORIGIN

Query Match 12.3%; Score 353.4; DB 6; Length 447;
Best Local Similarity 98.3%; Pred. No. 4.8e-69; Indels 0; Gaps 0;
Matches 357; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 ATGGCTCACTATATTACATTTCTCTGCATGGTTTTGGTGGCTCTTCTCAGAAATCTGTG 60
DB 85 ATGGCTCACTATATTACATTTCTCTGCATGGTTTTGGTGGCTCTTCTCAGAAATCTGTG 144
QY 61 TTAGCTGAAGATGGGGAAGTAAGATCAAGTTGTCTGCTACTGCTCGACAGATTTAGTTTC 120
DB 145 TTAGCTGAAGATGGGGAAGTAAGATCAAGTTGTCTGCTACTGCTCGACAGATTTAGTTTC 204
QY 121 ATCTTAGATGGCTCTTATAGTGTGGCCAGAAAACCTTTGAAATAGTGAAGAGTGGCTT 180
DB 205 ATCTTAGATGGCTCTTATAGTGTGGCCAGAAAACCTTTGAAATAGTGAAGAGTGGCTT 264
QY 181 GTCAATATCACAAAACCTTTGACATAGGGCCGAAGTTTATCAAGTTGAGTGGTCAA 240
DB 265 GTCAATATCACAAAACCTTTGACATAGGGCCGAAGTTTATCAAGTTGAGTGGTCAA 324
QY 241 TATAGTGAATACCTGCTGCTGAGATTCCTCTCGAAGCTATGATTCAGGAGACATTTG 300
DB 325 TATAGTGAATACCTGCTGCTGAGATTCCTCTCGAAGCTATGATTCAGGAGACATTTG 384
QY 301 ACGGAGCAGTGGAAATCCATCTTACTTAGGAGGAAACACAAAGACAGGAGGCCATC 360
DB 385 ACGGAGCAGTGGAAATCCATCTTACTTAGGAGGAAACACAAAGACAGGAGGCCATC 444
QY 361 CAG 363
DB 445 GAG 447

RESULT 13
CHKGSCND
LOCUS
DEFINITION
Chicken chondrocyte specific short collagen, encoding HMW/LMW precursor.
ACCESSION
K01702
VERSION
K01702.1 GI:211498
KEYWORDS
collagen; matrix protein.
SOURCE
Gallus gallus (chicken)
ORGANISM
Gallus gallus
REFERENCE
1. (bases 1 to 3164)
NINOMIYA,Y. and OLSEN,B.R.
Synthesis and characterization of cDNA encoding a
cartilage-specific short collagen
Proc. Natl. Acad. Sci. U.S.A. 81 (10), 3014-3018 (1984)
JOURNAL
MEDLINE
84221895
PUBMED
6328487
COMMENT
Original source text: Chicken embryo (17 days) sternal cartilage
and calvarial bone, cDNA to mRNA, clone pYN1738.
This mRNA, which codes for a collagenous polypeptide, has several
unusual characteristics. First, the total length of the molecule is
only about half that of pro-alpha-1(I) chains. Second, the
molecule contains three collagenous domains connected by short,
noncollagenous peptides. Third, several cysteinyl residues are
located within the noncollagenous peptides. In the sites table NC
denotes a noncollagenous region, COL denotes a collagenous region.
(i) reports that this mRNA encodes the polypeptide precursor of one
of the subunits of HMW and LMW collagens. They suggest that HMW

```

Db 27494 TTTCTGACTAGATAGCAGTGTACTTACCGATGGCAATCCCAAGATGACGTCAGAGATG 27553  
 Qy 461 CAGCTCAAGCAGCAAGATAGTAGATTAACATTTATTTGCTATTGCTGTTGCTCAGAAA 520  
 Db 27554 CAGCTCAAGCAGCAAGATAGTAGATTAACATTTATTTGCTATTGCTGTTGCTCAGAAA 27613  
 Qy 521 CAGAAGATCGGCAACTAGAGCTATTGCCAAGCAGCTTCGTTACTATTGCTTTATG 580  
 Db 27614 CAGAAGATCGGCAACTAGAGCTATTGCCAAGCAGCTTCGTTACTATTGCTTTATG 27673  
 Qy 581 TGGAGACTATATTGCAATATCCAAATTAAGGGAAGTATGAAGCAGAGAACTTTGTGAAG 640  
 Db 27674 TGGAGACTATATTGCAATATCCAAATTAAGGGAAGTATGAAGCAGAGAACTTTGTGAAG 27733

## RESULT 10

HS708F5/c

LOCUS

DEFINITION

Human DNA sequence from clone RP4-708F5 on chromosome 6p11.2-12.3,

complete sequence.

ACCESSION

AL031782

VERSION

AL031782.1 GI:4164360

KEYWORDS

HTG.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 126501)

Griffiths, C.

Direct Submission

Submitted (04-MAR-2003) Wellcome Trust Sanger Institute, Hinxton,

Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

On Jan 19, 1999 this sequence version replaced gi:3947843.

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: humquery@sanger.ac.uk

-----

During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with

only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all

regions were either double-stranded or sequenced with an alternate

chemistry or covered by high quality data (i.e., phred quality &gt;=

30); an attempt was made to resolve all sequencing problems, such

as compressions and repeats; all regions were covered by at least

one plasmid subclone or more than one M13 subclone; and the

assembly was confirmed by restriction digest, except on the rare

occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession

numbers given in the feature table with their source databases:

En.: EMBL; Sw.: SWISSPROT; Tr.: TREMBL; Wp.: WormPEP; Information

on the WormPEP database can be found at

<http://www.sanger.ac.uk/Projects/C.elegans/wormpep> This sequence

was generated from part of bacterial clone contigs of human

Chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping

Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/Chr6>

RP4-708F5 is from the library RPCI-4 constructed by the group of

Pieter de Jong. For further details see

<http://www.chori.org/bacpac/home.htm>

VECTOR: pCYPAC2.

Location/Qualifiers

1..126501

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

## FEATURES

source

/chromosome="6"  
 /map="p11.2-12.3"  
 /clone="RP4-708F5"  
 /clone\_lib="RPCI-4"  
 BASE COUNT 44:77 a 23609 c 21902 g 36813 t  
 ORIGIN

Query Match 19.3%; Score 553.6; DB 9; Length 126501;  
 Best Local Similarity 95.2%; Pred. No. 4.9e-114;  
 Matches 57; Conservative 0; Mismatches 29; Incels 0; Gaps 0;

Qy 41 TGCCTTTTCAGAAATCTGTGTAGCTGAAGATGGGAAGTAAAGATCAAGTTGTCCTACTG 100  
 Db 50175 TGCCTTTTCAGAAATCTGTGTAGCTGAAGATGGGAAGTAAAGATCAAGTTGTCCTACTG 50116  
 Qy 101 CTCGACAGATTTAGTTTTCATCTTAGATGGCTCTTATAGTTGGTGGCCAGAAAACCTTTG 160  
 Db 50115 CTCGACAGATTTAGTTTTCATCTTAGATGGCTCTTATAGTTGGTGGCCAGAAAACCTTTG 50056  
 Qy 161 AATAGTGAAGAAGTGGCTTTGCAATATACAAAAAACTTTGACATAGGCGCCGAGATTTA 220  
 Db 50055 AATAGTGAAGAAGTGGCTTTGCAATATACAAAAAACTTTGACATAGGCGCCGAGATTTA 49996  
 Qy 221 TTCAAGTTGGAGTGGTTCATATAGTACTACCTGTGCTGGAGATTCCTCTCGGAAGCT 280  
 Db 49995 TTCAAGTTGGAGTGGTTCATATAGTACTACCTGTGCTGGAGATTCCTCTCGGAAGCT 49936  
 Qy 281 ATGATTTCAGGAACAATTTGACGCGAGTGGATCCATCTACTCTACTCTAGGAGAAACA 340  
 Db 49935 ATGATTTCAGGAACAATTTGACGCGAGTGGATCCATCTACTCTACTCTAGGAGAAACA 49876  
 Qy 341 CAAAGACAGGAAGGCCATCCAGTTTGGCTCGATTACCTTTTGGCAAGTCTCTCAGAT 400  
 Db 49875 CAAAGACAGGAAGGCCATCCAGTTTGGCTCGATTACCTTTTGGCAAGTCTCTCAGAT 49816  
 Qy 401 TTCTGACTAAGATAGCAGTGGTACTTAGGATGGCAATCCAGATGACGTCAGGATG 460  
 Db 49815 TTCTGACTAAGATAGCAGTGGTACTTAGGATGGCAATCCAGATGACGTCAGGATG 49756  
 Qy 461 CAGCTCAAGCAGCAAGAGATAGTAAAGATTTATTTGCTATTGCTATTGGTTTGGTTTCAAAA 520  
 Db 49755 CAGCTCAAGCAGCAAGAGATAGTAAAGATTTATTTGCTATTGGTTTGGTTTCAAAA 49696  
 Qy 521 CAGAAGATCGCAACTTAGAGCTATTGGCAAGCAGCTTCGTTACTATTGTTTATG 580  
 Db 49695 CAGAAGATCGCAACTTAGAGCTATTGGCAAGCAGCTTCGTTACTATTGTTTATG 49636  
 Qy 581 TGGAGACTATATTGCAATATCCAAATTAAGGGAAGTATGAAGCAGAGAACTTTGTGAAG 640  
 Db 49635 TGGAGACTATATTGCAATATCCAAATTAAGGGAAGTATGAAGCAGAGAACTTTGTGAAG 49576

## RESULT 11

AX721209

LOCUS

DEFINITION

Sequence 169 from Patent WO0220754.

ACCESSION

AX721209

VERSION

AX721209.1 GI:30422045

KEYWORDS

Homo sapiens (human)

SOURCE

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1

Lincoln, S.E., Alcus, C.M., Dufour, G.E., Chalup, M.S., Hillman, J.L.,

Jones, A., Yu, C.Y., Wright, R.J., Gietzen, D., Liu, Tomy, F., Yap, P.,

Bahl, C.R., Moniyara, M., Bradley, D., Rohatgi, S., Harris, B.,

Roseberry, A.M., Gerstin, E.H., Petalta, C.H., David, M., Panzer, S.,

Flores, V., Daffo, A., Marwaha, R., Chen, A., Chang, S.C. and Immar, R.R.

Molecules for diagnostics and therapeutics

Patent: WO 0220754-A 169 14-MAR-2002;

Incyte Genomics, Inc. (US)

Location/Qualifiers

Qy	481	AGTAAGATAACAATTATTTGCTATTGGTGTGGTTTCAGAAACAGAGAATGCCGAACCTTAGA	540
Db	676	AGTAAGATAACAATTATTTGCTATTGGTGTGGTTTCAGAAACAGAGAATGCCGAACCTTAGA	735
Qy	541	GCTATTGCCAACCAAGCCTTCGCTCTACTATTATGCTGTTTTATGTGGAAGACTATATTGCAATA	600
Db	736	GCTATTGCCAACCAAGCCTTCGCTCTACTATTATGCTGTTTTATGNGGAGACTATATTGCAATA	795
Qy	601	TCCAAATAAAGGGAAGTGATGAAGCAGAAACTTTGTGAAGAACTCTGCTCTCCAACACGA	660
Db	796	TCCAAATAAAGGGAAGTGATGAAGCAGAAACTTTGTGAAGAACTCTGCTCTCCAACACGA	855
Qy	661	ATTCCAGTGCACGCTCGTGATGAAGGGGATTTGATATCTTTTAGGTTAGATGTAAAT	720
Db	856	ATTCCAGTGCACGCTCGTGATGAAGGGGATTTGATATTC-TTTAGGTTTAGATGTAAAT	914
Qy	721	AAAAAGGTTAAGAAAGAAATACAGCTTTCACCAAAAAAGATAAAGAGATGAAGTAACA	780
Db	915	AAAAAGGTTAAGAAAGAAATACAGCTTTCACCAAAAAAGATAAAGAGATGAAGTAACA	974
Qy	781	TCAAAAGTTGATTTATCAGAACTTCACAGCAATGTTTTCCAGAAAGTCTTCTCCATCA	840
Db	975	TCAAAAGTTGATTTATCAGAACTTCACAGCAATGTTTTCCAGAAAGTCTTCTCCATCA	1034
Qy	841	TATGTAATTTGCTCTACTCAAGATTTAAAGTCAAGAAAAATTTGGGATTTATGGAGAAATA	900
Db	1035	TATGTAATTTGCTCTACTCAAGATTTAAAGTCAAGAAAAATTTGGGATTTATGGAGAAATA	1094
Qy	901	TTAACTATTGATGGAGGCCA 921	
Db	1095	TTAACTATTGATGGAGGCCA 1115	

RESULT 9  
 AF438327.2  
 WPCOMENT

Sequence split into 4 fragments    LOCUS AF438327    Accession AF438327  
 Fragment Name    Begin    End  
 AF438327.0    1    110000  
 AF438327.1    100001    210000  
 AF438327.2    200001    310000  
 AF438327.3    300001    381696

Continuation (3 of 4) of AF438327 from base 200001 (AF438327 Homo sapiens alpha

Query Match	19.3%	Score 553.6;	DB 9;	Length 110000;
Best Local Similarity	95.2%	Pred. No. 4.8e-114;		
Matches	571;	Conservative	0;	Mismatches 29;
			Indels	0;
			Gaps	0;

Qy	41	TCGCTTCTTCAGAACTCTGCTGTTAGCTGAAGATGGGGAAGTAAAGTCAAGTTCTGCTACTG	100
Db	27134	TCGCTTCTTCAGAACTCTGCTGTTAGCTGAAGATGGGGAAGTAAAGTCAAGTTCTGCTACTG	100
Qy	101	CTCCGACAGATTTAGTATTTCATCTTAGATGGCTCTTATAGTGTGTGGCCAGAAAACCTTTG	160
Db	27194	CTCCGACAGATTTAGTATTTCATCTTAGATGGCTCTTATAGTGTGTGGCCAGAAAACCTTTG	2725
Qy	161	AAATAGTGAAGAAAGTGGCTTGTCAATATCAAAAAACCTTTGACATAGGCCCGAAGTTTA	220
Db	27254	AAATAGTGAAGAAAGTGGCTTGTCAATATCAAAAAACCTTTGACATAGGCCCGAAGTTTA	2731
Qy	221	TTCAAGTTGCAGTGGTTTCAATATAGTACGTACACCTGTGCTGGAGATTCCTCTCGGAAGCT	280
Db	27314	TTCAAGTTGCAGTGGTTTCAATATAGTACGTACACCTGTGCTGGAGATTCCTCTCGGAAGCT	2737
Qy	281	ATGATTCAGGAGAACATTTGACGGCAGCAGTGGAAATCCATCTACTTAGGAGGAAACA	340
Db	27374	ATGATTCAGGAGAACATTTGACGGCAGCAGTGGAAATCCATCTACTTAGGAGGAAACA	2743
Qy	341	CRAGACAGGGAAGGCCATCCAGTTTGCCTCGATTACCTTTTTCGAAGTCTCTCAGAT	400
Db	27434	CRAGACAGGGAAGGCCATCCAGTTTGCCTCGATTACCTTTTTCGAAGTCTCTCAGAT	2749
Qy	401	TTCTGACTAAGATAGCAGTGSTACTTTACGGAATGGCAAAATCCCAAGATGACGTCAAGGATG	460





TITLE Direct Submission  
JOURNAL Submitted (04-JUN-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7 Kazusa-Kanetani, Kisarazu, Chiba 292-0812, Japan (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986) NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- and 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.

FEATURES  
source  
1..2295  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="NTONG2007034"  
/tissue\_type="tongue"  
/clone\_lib="NTONG2"  
/note="cloning vector: pME18SFL3"

BASE COUNT 683 a 427 c 565 g 620 t

ORIGIN  
Query Match 36.6%; Score 1048; DB 9; Length 2295;  
Best Local Similarity 99.8%; Pred. No. 3.4e-226;  
Matches 1060; Conservative 0; Mismatches 0; Indels 2; Gaps 1;

QY 1804 GGAATCCCAAGGATTTCTGGAAACCGAGGATTAATGGSCCAAAAGGGAGAAATTTGGCCCT 1863  
DB 169 GGAATCCCAAGGATTTCTGGAAACCGAGGATTAATGGSCCAAAAGGGAGAAATTTGGCCCT 228  
QY 1864 CCAGACAGCAAGGAAAAAAGAGCCCGCAGGATGCTTGTATTAATGGGAAGCAATGGC 1923  
DB 229 CCAGACAGCAAGGAAAAAAGAGCCCGCAGGATGCTTGTATTAATGGGAAGCAATGGC 288  
QY 1924 TCACAGGACAGCTTGAACACCGGGATCTAAGGAGCAAGGATGAACCTTGAATTCAA 1983  
DB 289 TCACAGGACAGCTTGAACACCGGGATCTAAGGAGCAAGGATGAACCTTGAATTCAA 345  
QY 1984 GGGATGCTGGGCTTCTGGGCTCAAGGGAGAAACAGGAGCAACGGGTTCCCGAGGAGAA 2043  
DB 349 GGGATGCTGGGCTTCTGGGCTCAAGGGAGAAACAGGAGCAACGGGTTCCCGAGGAGAA 408  
QY 2044 CCAGATACATGGGTTTACCCGGGATTAAGGAAAAAAGGGGAGCAAGGAAATCAAGGT 2103  
DB 409 CCAGATACATGGGTTTACCCGGGATTAAGGAAAAAAGGGGAGCAAGGAAATCAAGGT 468  
QY 2104 GAAAAGGTATTTCAGGTTCAAAAGGGAGAAATGAAGACAGGGAAATTCAGGGCAACAG 2163  
DB 469 GAAAAGGTATTTCAGGTTCAAAAGGGAGAAATGAAGACAGGGAAATTCAGGGCAACAG 528  
QY 2164 GAAAATTCAGGCAATCATGTTGCAAAAGGAGAGAGAGGTGAAAAGGGAGAAACCTGGTGTG 2223  
DB 529 GAAAATTCAGGCAATCATGTTGCAAAAG--GAGAGAGGTGAAAAGGGAGAAACCTGGTGTG 586  
QY 2224 CCAGGTGCCATTCAGATCAAAAGGAGATCTGGGGTGGATGCTGATGGGCGCCGAGGT 2283  
DB 587 CCAGGTGCCATTCAGATCAAAAGGAGATCTGGGGTGGATGCTGATGGGCGCCGAGGT 646  
QY 2284 CTTAAGGGCAACCTGGGATCCAGGTCTCAGGGACCCCGCAGGTTTGGATGGGAAGCCC 2343  
DB 647 CTTAAGGGCAACCTGGGATCCAGGTCTCAGGGACCCCGCAGGTTTGGATGGGAAGCCC 706  
QY 2344 GGAAGAGATTTTCAGAAATTTATTCAGAGTTTCAGAGTTTCAGAGTTTCAGAGTTTCAG 2403  
DB 707 GGAAGAGATTTTCAGAAATTTATTCAGAGTTTCAGAGTTTCAGAGTTTCAGAGTTTCAG 766  
QY 2404 CTACCACTCTTACTTCAGAGTTTCAGAAATTTAGAAATTTGTGATCATTTGCTCTCCCAACAT 2463  
DB 767 CTACCACTCTTACTTCAGAGTTTCAGAAATTTAGAAATTTGTGATCATTTGCTCTCCCAACAT 826  
QY 2464 GGCTCCCCGGGTATTCCTGGGCGCA CTTGGTCCGATAGGCCCGAGGGTCCCGAGAGGATTA 2523

0  
DB 327 GGCTCCCGGGTATTTCTGGGCACTGGTCCGATAGGCCACAGAGAGTCCCGAGGATTA 886  
QY 2524 CTTGTTTCCAGGAGAGATGGTTCCTCGATAGTGGTGTCTCTGACGTCCTGACGTCAGGT 2583  
DB 887 CTTGTTTCCAGGAGAGATGGTTCCTCGATAGTGGTGTCTCTGACGTCCTGACGTCAGGT 946  
QY 2584 GTCACAGGATTTAAAGGCTTACAGGAGAAATCGGGAAGAAAGGAGGAGGAGGTTGGG 2643  
DB 947 GTCACAGGATTTAAAGGCTTACAGGAGAAATCGGGAAGAAAGGAGGAGGAGGTTGGG 1006  
QY 2644 TATCTCGAGAACAGGTCCTCTGGTCCCCCAGGTCCAGAGGGCCCTCTTGAATTAAGC 2703  
DB 1007 TATCTCGAGAACAGGTCCTCTGGTCCCCCAGGTCCAGAGGGCCCTCTTGAATTAAGC 1066  
QY 2704 AAAGAGGTCCTCCAGGAGACCCAGGTCTCCTGGCAAGATGAGACCATGGAACCT 2763  
DB 1067 AAAGAGGTCCTCCAGGAGACCCAGGTCTCCTGGCAAGATGAGACCATGGAACCT 1126  
QY 2764 GGAATCCCAAGGCAACAGGCCCCCAGGATCTCGACACCATCACTATGTTTAGTGTA 2823  
DB 1127 GGAATCCCAAGGCAACAGGCCCCCAGGATCTCGACACCATCACTATGTTTAGTGTA 1186  
QY 2824 ATTGCCAGAGAGATCCGTTCCAGAAAGGACCAAACTATTAG 2865  
DB 1187 ATTGCCAGAGAGATCCGTTCCAGAAAGGACCAAACTATTAG 1228

RESULT 6  
AX43041C AX43041C 1184 bp DNA linear PAT 28-JUN-2002  
LOCUS Sequence 192 from Patent WO0240715.  
DEFINITION ACCESSION AX43041C  
VERSION AX430410.1 GI:21655774  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1  
AUTHORS Chalup M.S., Altus C.W., Lincoln S.E., Dufour G.E. and Jackson S.  
TITLE Molecules for disease detection and treatment  
JOURNAL Patent: WO 0240715-A 192 23-MAY-2002;  
INCITE GENOMICS INC (US)  
FEATURES  
source  
1..1184  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/note="Incite ID No: LJ:108819.1:2000SEP08"

BASE COUNT 368 a 227 c 272 g 327 t

ORIGIN  
Query Match 34.0%; Score 974; DB 6; Length 1184;  
Best Local Similarity 100.0%; Pred. No. 1.8e-209;  
Matches 974; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCTCACTATATTACATTTCTCTGCATGGTTTGGTGTCTCTCAGAAATTCGTGTG 60  
DB 211 ATGGCTCACTATATTACATTTCTCTGCATGGTTTGGTGTCTCTCAGAAATTCGTGTG 276  
QY 61 TTAGCTGAAGATGGGGAAGTAAGATCAAGTTGTCGTACTGCTCCGACAGATTTAGTTTTC 120  
DB 271 TTAGCTGAAGATGGGGAAGTAAGATCAAGTTGTCGTACTGCTCCGACAGATTTAGTTTTC 330  
QY 121 ATCTTAGATGGCTCTTATAGTGTGGGCCCGAAAAACCTTGAATAGTGAAGAGTGGCTT 180  
DB 331 ATCTTAGATGGCTCTTATAGTGTGGGCCCGAAAAACCTTGAATAGTGAAGAGTGGCTT 390  
QY 181 GTCATATACAAAAAATTTGATAGGCGCGAGTTTATTCAGTTGAGTGGTTCATCA 240  
DB 391 GTCATATACAAAAAATTTGATAGGCGCGAGTTTATTCAGTTGAGTGGTTCATCA 450



QY	1081	GTGACTTTGTATTTATGATGACCAACAAATTTGAAACAAAGCCCTTACATCCAGTTTATAGGG	1140	QY	2152	CCAGGGCAACAGGGAATTTCAAGGCCATCTGGTGCRAAAGAGAGAGAGAGTGAAGGGA	2211
DB	1283	GTGACTTTGTATTTATGATGACCAACAAATTTGAAACAAAGCCCTTACATCCAGTTTATAGGG	1342	DB	2363	CCAGGGCAACAGGGAATTTCAAGGCCATCTGGTGCRAAAGAGAGAGAGTGAAGGGA	2422
QY	1141	ATCTTGATCAATGGGCAACCAATTTGGAATATTTCTGGAAAGAGAACTGTTTCAG	1200	QY	2212	GAACTCGTGTCCGAGGTGCCATTTGGATCAAAAGAGAACTCTGGGGTGGATGGCTTGATG	2271
DB	1343	ATCTTGATCAATGGGCAACCAATTTGGAATATTTCTGGAAAGAGAACTGTTTCAG	1402	DB	2423	GAACTCGTGTCCGAGGTGCCATTTGGATCAAAAGAGAACTCTGGGGTGGATGGCTTGATG	2482
QY	1201	TTTGATGTCCAAAGTTGGAAATCTACTGTGACCAAGAACAAACCCGGGAGACGCA	1260	QY	2272	GGCCCGGAGGTCTTAAGGGCAACCTGGGGATCAAGGTCTCAGGACCCCGAGGTTG	2331
DB	1403	TTTGATGTCCAAAGTTGGAAATCTACTGTGACCAAGAACAAACCCGGGAGACGCA	1462	DB	2483	GGCCCGGAGGTCTTAAGGGCAACCTGGGGATCAAGGTCTCAGGACCCCGAGGTTG	2542
QY	1261	TCTGAGATTCCTGGATTT-----TGCCTTAATGTCCTGAGTGTAGGTTCAACT	1311	QY	2332	GATGGGAAGCCCGGAAGAGAGTTTTCAGAACAAATTTATCGACAAGTTTCACAGATGTA	2391
DB	1463	TCTGAGATTCCTGGATTT-----TGCCTTAATGTCCTGAGTGTAGGTTCAACT	1522	DB	2543	GATGGGAAGCCCGGAAGAGAGTTTTCAGAACAAATTTATCGACAAGTTTCACAGATGTA	2602
QY	1312	CCAGCTCCCTGTATTTGTCTCCGGGAAACCCAGGACTTCAAGGCCCCCAAGGTGACCCCT	1371	QY	2392	ATAAGAGCCCACTACCAAGTCTTACTTCAGAGTGGAGAAATTAAGAAATTTGATCATGTC	2451
DB	1523	CCAGCTCCCTGTATTTGTCTCCGGGAAACCCAGGACTTCAAGGCCCCCAAGGTGACCCCT	1582	DB	2603	ATAAGAGCCCACTACCAAGTCTTACTTCAGAGTGGAGAAATTAAGAAATTTGATCATGTC	2662
QY	1372	GGACTGCTGGGAACCTGGTACCTCGAGAACCTGGTCAAGATGGTAAAGCTGGATAT	1431	QY	2452	CTGTCCCAACATGGCTCCCGGGTATTTCTGGGCCACCTGTCCGATAGGCCAGAGGGT	2511
DB	1583	GGACTGCTGGGAACCTGGTACCTCGAGAACCTGGTCAAGATGGTAAAGCTGGATAT	1642	DB	2663	CTGTCCCAACATGGCTCCCGGGTATTTCTGGGCCACCTGTCCGATAGGCCAGAGGGT	2722
QY	1432	CAGGGAATTCAGGGACACAGGTGTTCCAGGATCTCCAGGAATACAAAGGAGCTCGAGGA	1491	QY	2512	CCAGAGGATTACCTGGTTTCCAGAGAGAGATGGTGTCTCTGATTTAGTGGGTGTCCT	2571
DB	1643	CAGGGAATTCAGGGACACAGGTGTTCCAGGATCTCCAGGAATACAAAGGAGCTCGAGGA	1702	DB	2723	CCAGAGGATTACCTGGTTTCCAGAGAGAGATGGTGTCTCTGATTTAGTGGGTGTCCT	2782
QY	1492	CTACAGGTTACAAAGGAAACAGGGCGAGATGGTGACAGGGTGATCGTGATTCCT	1551	QY	2572	GSAGTCCAGGTGTCCAGAGGATTAAGAGCCCTACAGAGAGAAATGGGAAAAGGAGC	2631
DB	1703	CTACAGGTTACAAAGGAAACAGGGCGAGATGGTGACAGGGTGATCGTGATTCCT	1762	DB	2783	GSAGTCCAGGTGTCCAGAGGATTAAGAGCCCTACAGAGAGAAATGGGAAAAGGAGC	2842
QY	1552	GGTTTCTCGGCTTCATGCGATGCCAGATCAAGAGGTGAAATCGGTGGCCCAAGGAGAC	1611	QY	2632	CAAGGTTTGGGTAATCTCGAGAACAGGTCTCTGTGTCCTCCCGAGGTCCAGAGGGCCCT	2691
DB	1763	GGTTTCTCGGCTTCATGCGATGCCAGATCAAGAGGTGAAATCGGTGGCCCAAGGAGAC	1822	DB	2843	CAAGGTTTGGGTAATCTCGAGAACAGGTCTCTGTGTCCTCCCGAGGTCCAGAGGGCCCT	2902
QY	1612	AAAGATCACTGGATTTTATGGCAAAAGAGGTGCAAAAGGTGAAAAGGGGAATGCTGGC	1671	QY	2692	CTCGAATTAAGCAAAAGAGGTCTCTCGAGAGACCCAGGTCTCTCTGGCAAGATGGAGAC	2751
DB	1823	AAAGATCACTGGATTTTATGGCAAAAGAGGTGCAAAAGGTGAAAAGGGGAATGCTGGC	1882	DB	2903	CTCGAATTAAGCAAAAGAGGTCTCTCGAGAGACCCAGGTCTCTCTGGCAAGATGGAGAC	2962
QY	1672	TTCCCTGGCTCCCTGGACTCTCGAGAACCCAGAGACATCGGAAAGATGGATTAATG	1731	QY	2752	CATGGAACCTCGGAATCAAGGGCAACAGGCCCCCGAGGCATCTCGACCCATCACTA	2811
DB	1883	TTCCCTGGCTCCCTGGACTCTCGAGAACCCAGAGACATCGGAAAGATGGATTAATG	1942	DB	2963	CATGGAACCTCGGAATCAAGGGCAACAGGCCCCCGAGGCATCTCGACCCATCACTA	3022
QY	1732	GGTAGTCCGGTTCAAGGGAGAGCAGGATCCCTGTGTCTCCGGGCGAGGATGGAA CA	1791	QY	2812	TGTTTGTAGTGAATTTCCAGAGAGATCCGTTCAGAAAGAGCCAAACTATTAG	2865
DB	1943	GGTAGTCCGGTTCAAGGGAGAGCAGGATCCCTGTGTCTCCGGGCGAGGATGGAA CA	2002	DB	3023	TGTTTGTAGTGAATTTCCAGAGAGATCCGTTCAGAAAGAGCCAAACTATTAG	3076
QY	1792	CGGGGAGAGCTGGATCCAGGATTTCTCGGAACCCAGGATTAATGGCCCAAGGGA	1851	RESULT 5			
DB	2003	CGGGGAGAGCTGGATCCAGGATTTCTCGGAACCCAGGATTAATGGCCCAAGGGA	2062	AK096444	2295 bp	mRNA	linear
QY	1852	GAAATTTGGGCTCCAGGACAGCAAGGAAAAAGGAGCCCGAGGATGCTGTTTATG	1911	LOCUS	AK096444		
DB	2063	GAAATTTGGGCTCCAGGACAGCAAGGAAAAAGGAGCCCGAGGATGCTGTTTATG	2122	DEFINITION	Homo sapiens cDNA FLJ39125 fis, clone NTONG2007034.		
QY	1912	GGAAAGTAATGGCTCACCGGCGAGCTGGAAACCCGGATCTAAGGGAGCAAGGTGAA	1971	ACCESSION	AK096444		
DB	2123	GGAAAGTAATGGCTCACCGGCGAGCTGGAAACCCGGATCTAAGGGAGCAAGGTGAA	2182	VERSION	AK096444.1	GI:21755943	
QY	1972	CCTGGATTCAGGGATGCTGGGGTCTCTGGGCTCAAGGGAGAACCCAGGAGCAAGGGT	2031	KEYWORDS	oligo capping; fis (full insert sequence).		
DB	2183	CCTGGATTCAGGGATGCTGGGGTCTCTGGGCTCAAGGGAGAACCCAGGAGCAAGGGT	2242	SOURCE	Homo sapiens (human)		
QY	2032	TCCCGAGGAAACAGGATACATGGTTTACCCGGGATTCAGGAAAAAGGGGAGCAAA	2091	ORGANISM	Homo sapiens		
DB	2243	TCCCGAGGAAACAGGATACATGGTTTACCCGGGATTCAGGAAAAAGGGGAGCAAA	2322	REFERENCE	1		
QY	2092	GGAAATCAAGGTGAAAGGATTTACGGGTCAAGGGAGAAATTAAGGACAGGGAAT	2151	AUTHORS	Oghira, A., Takahashi-Fujii, A., Tarase, T., Imose, N., Takeuchi, K.,		
DB	2303	GGAAATCAAGGTGAAAGGATTTACGGGTCAAGGGAGAAATTAAGGACAGGGAAT	2362		Zita, M., Musashino, K., Yuchi, H., Hara, H., Sugiyama, T., Irie, R.,		
					Ctsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y.,		
					Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yakashita, H.,		
					Matsumoto, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K.,		
					Wagatsuru, M., Murakawa, K., Kanehori, K., Sugiyama, A., Kawakami, B.,		
					Suzuki, Y., Sugano, S., Nagahara, K., Masuho, Y., Nagai, K. and		
					Isogai, T.		
					KEDO human cDNA sequencing project		
					Unpublished		
					REFERENCE 2 (bases 1 to 2295)		
					Isogai, T. and Yamamoto, J.		

LOCUS HSM801597 4160 bp mRNA linear PRI 20-MAR-2002  
DEFINITION Homo sapiens mRNA; cDNA DKFZp564B052 (from clone DKFZp564B052);  
complete cds.

ACCESSION AL136624

VERSION AL136624.1 GI:12052773

KEYWORDS

SOURCE Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 4160)

Duesterhoeft A., Lauber J., Mewes H.W., Gassenhuber J. and

Wienann S.

TITLE Direct Submission

JOURNAL Submitted (12-MAR-2002) MIPS, Am Klopferspitz 18a, D-82152

COMMENT

Cloned from S. Wienann, Molecular Genome Analysis, German Cancer

Research Center (DKFZ); Email s.wienann@dkfz-heidelberg.de;

sequenced by Qiagen (Hilden/Germany) within the cDNA sequencing

consortium of the German Genome Project.

This clone (DKFZp564B052) is available at the RZPD in Berlin.

Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059

Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further

information about the clone and the sequencing project is available

at <http://mips.gsf.de/proj/cDNA/>.

FEATURES

source

Location/Qualifiers

1..4160

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/map="6"

/clone="DKFZp564B052"

/tissue\_type="brain"

/clone\_lib="564 (synonym: hfbz2). Vector pAMP1: host

X1-2blue; sites NotI + SalI"

/dev\_stage="fetal"

1..4160

/gene="DKFZp564B052"

/gene="DKFZp564B052"

/note="DJ708F5.1, differences to genmodel, similarity to

collagen proteins"

/codon\_start=1

/product="hypothetical protein"

/protein\_id="CA86559.1"

/db\_xref="GI:12052774"

/db\_xref="SPTREMBL:Q9H0V3"

/translation="MAHYITFLCMVLLVLLQNSVLAEDGEVRSRCRTAPDLVFLDGG

SYSGPENFEIVKMLNITKNFDIGPKFIOGVVQYSDYVLEIPIUGSDGEHLTA

AVESTLYLGGNTCKATQFALDYLDKSSRFLTKIAVLTDKGSDVDKDAQAARD

SKITLFAIGVSETEDELAIANKPSTVYVYEDVIALSKIREVMKOLCEESVCF

TRIPVAARDERGFIDLLGVNKKVKKRIOLSPKIKGYEVSQVLDSELTSNVFPEG

LPSPVQVTRQFVKKKIWDLRILITDGRFQIAVTLNGVDKILLFTTSVINGSCVW

TRANPQVILFDEGMHQIRLIVTEQDVLTIIDDOQIENKPLHPVLGILLNGTQIGKY

SGKETVIGVQKRIYCDPEQNNRETACEIPGNGSLNPSDVGSTPAPCIQPGK

PGQKFGDGLGPNMPYQPGQDQKPGYQIAGTGVPGSPGQIQAARGLPGYKGP

GRDGKGRGLPGFLHGMPSKSGEMKAGDKGSPGYGKGAAGKGNAGPGLPG

PAGEPGRKGDGLMGS PGFKGAGS PGAPGQDGTGPPIPGPNRGLMGKGEIGP

PGQKKAQKMPGLMGSNGSPGPGTSGSKSGEPGIQMPGASGLKCEPGATGSP

CEPYGGLPLPGKDKGNQGEKGIQKQKNGRQGI PQQGLIQHGHAKGERGKG

EPGVKGLSGSGESVDGLMGPAGFGQPGDFGPGFGLGKRFGRFSQFIQVCT

DVIRAQVLQSGRINRCHLSQHS GPFGPGPIGEPGRGLPGLPGRDGVPLG

VGVPRPVRGLGLKGLPGRNGKSGSGPGQGPQPGPQPGPGPGPGPGPGPGPG

PKGDGHKPGIQCGQPGPGICDPSLCSFVSIARRDPPRKGPNY"

4120..4125

/polyA\_signal

/gene="DKFZp564B052"

4141

/polyA\_site

/gene="DKFZp564B052"

BASE COUNT 1271 a 795 c 1026 g 1068 t

ORIGIN

Query Match

Best Local Similarity

99.1%; Score 2839.6; DB 9; Length 4160;

99.5%; Pred. No. 0;

Matches	2861;	Conservative	0;	Mismatches	4;	Indels	9;	Gaps	1;
QY	1	ATGGCTCACTATATTACATTTCTCTGCATGGTTTTGGTGGTCTTCTTTCAGAAATCTGTG	60						
DB	203	ATGGCTCACTATATTACATTTCTCTGCATGGTTTTGGTGGTCTTCTTTCAGAAATCTGTG	262						
QY	61	TTAGTGAAGTGGGGAAGTAAATCAAGTTGCTGCTCCGACAGATTTAGTTTTTC	120						
DB	263	TTAGTGAAGTGGGGAAGTAAATCAAGTTGCTGCTCCGACAGATTTAGTTTTTC	322						
QY	121	ATCTTAGATGCTCTTATAGTGTGGCCAGAAAACTTTGAAATAGTGAAGAAGTGGCTT	180						
DB	323	ATCTTAGATGCTCTTATAGTGTGGCCAGAAAACTTTGAAATAGTGAAGAAGTGGCTT	382						
QY	181	GTCAATATCAAAAAAATTTTGACATAGGSCCAAGTTTATTCAAGTTGAGTGGTTCAA	240						
DB	383	GTCAATATCAAAAAAATTTTGACATAGGSCCAAGTTTATTCAAGTTGAGTGGTTCAA	442						
QY	241	TATAGTCACTACCTGTGCTGGAGATTCCTCTCGGAAGCTATGATTGAGGAGAACTTTG	300						
DB	443	TATAGTCACTACCTGTGCTGGAGATTCCTCTCGGAAGCTATGATTGAGGAGAACTTTG	502						
QY	301	ACGCACAGTGGAAATCCATCTCTACTTAGGAGGAAACACAAAGACAGGAGGCAATC	360						
DB	503	ACGCACAGTGGAAATCCATCTCTACTTAGGAGGAAACACAAAGACAGGAGGCAATC	562						
QY	361	CAGTTTGGCTGCGATTACCTTTTGGCAAGTCTCTCAGATTCTGACTAAGATAGCAGTG	420						
DB	563	CAGTTTGGCTGCGATTACCTTTTGGCAAGTCTCTCAGATTCTGACTAAGATAGCAGTG	622						
QY	421	GTACTACGATGGCAATCCCAAGTACGTTCAAGATGCACTCAACACGACAGAT	480						
DB	623	GTACTACGATGGCAATCCCAAGTACGTTCAAGATGCACTCAACACGACAGAT	682						
QY	481	AGTAAGATAACATTTATTGCTATTGTTGGTTCAGAAAAACAGAGATGCCAACTTAGA	540						
DB	683	AGTAAGATAACATTTATTGCTATTGTTGGTTCAGAAAAACAGAGATGCCAACTTAGA	742						
QY	541	GCTATTGCCAAACAGCTTCGTTACTATTGTTGTTTATGTTGAAAGTATATGCAATA	600						
DB	743	GCTATTGCCAAACAGCTTCGTTACTATTGTTGTTTATGTTGAAAGTATATGTTGCAATA	802						
QY	601	TCCAAATATAGGGAAGTATGTAAGACAGAACTTTGTGAGAAATCTGCTGTCACACCA	660						
DB	803	TCCAAATATAGGGAAGTATGTAAGACAGAACTTTGTGAGAAATCTGCTGTCACACCA	862						
QY	661	ATTCAGTGGCAGCTCTGATGAAAGGGGATTTGATATTCTTTTAGGTTTAGATGTAAT	720						
DB	863	ATTCAGTGGCAGCTCTGATGAAAGGGGATTTGATATTCTTTTGGGTTTAGATGTAAT	922						
QY	721	AAAAGGTTAAGAAACAAATACAGCTTTACCAAAAAAGATAAAGGATATGAAGTAACA	780						
DB	923	AAAAGGTTAAGAAACAAATACAGCTTTACCAAAAAAGATAAAGGATATGAAGTAACA	982						
QY	781	TCAAAGTTGATTTATCAGAACTCACAAGCAATGTTTTCCCAAGAGGCTTCTCTCATCA	840						
DB	983	TCAAAGTTGATTTATCAGAACTCACAAGCAATGTTTTCCCAAGAGGCTTCTCTCATCA	1042						
QY	841	TATGTTTGTGCTCTACTCAAGATTTTAAAGTCAAGAAATTTGGGATTTAGGAGAATA	900						
DB	1043	TATGTTTGTGCTCTACTCAAGATTTTAAAGTCAAGAAATTTGGGATTTAGGAGAATA	1102						
QY	901	TTAACTATTGAGGAGGCCACAATAAGCAGTTACCTTTAAATGGTGGACAAAAATCTTA	960						
DB	1103	TTAACTATTGAGGAGGCCACAATAAGCAGTTACCTTTAAATGGTGGACAAAAATCTTA	1162						
QY	961	TTATTACAAACACAGCGTAATTAATGCTCACAAGTGGTTACCTTTGCTAACCTCAA	1020						
DB	1263	TTATTACAAACACAGCGTAATTAATGCTCACAAGTGGTTACCTTTGCTAACCTCAA	1222						
QY	1021	GTTAAGAGCTGTTTGTGATGAAGGCTGGCACCAAATTCGCTCTTCTTGTAGTAAACAGAA	1080						
DB	1223	GTTAAGAGCTGTTTGTGATGAAGGCTGGCACCAAATTCGCTCTTCTTGTAGTAAACAGAA	1282						

|||||  
923 AAAAAAGGTTAAGAAAAGATACAGCTTTTCA CAAAAAAGATAAAGGATATCAAGATACA 982  
Qy  
781 TCAAAAGTTGATTTATCAGAACTCAAGCAATGTTTTCCAGAGAGTCTTCTCCATCA 840  
Db  
983 TCAAAAGTTGATTTATCAGAACTCAAGCAATGTTTTCCAGAGAGTCTTCTCCATCA 1042  
Qy  
841 TATGATTTGCTGCTACTCAAGATTTAAGTCAAGAAATTTGGATTTATGAGAA 900  
Db  
1043 TATGATTTGCTGCTACTCAAGATTTAAGTCAAGAAATTTGGATTTATGAGAA 1102  
Qy  
901 TTAACATTATGATGGAAGGCCAATAATAGCAGTTACTTAATGTTGTCGACAAATCTTA 960  
Db  
1103 TTAACATTATGATGGAAGGCCAATAATAGCAGTTACTTAATGTTGTCGACAAATCTTA 1162  
Qy  
961 TTAATTTACAAACAGCGTAAATTAATGCTCAAGTGGTTACTTTTGTCTAAACCTCAA 1020  
Db  
1163 TTAATTTACAAACAGCGTAAATTAATGCTCAAGTGGTTACTTTTGTCTAAACCTCAA 1222  
Qy  
1021 GTTAAGACGTTGTTGATCAAGCGCTGACCAATTCGTTCTTGTAGTAACAGAAAGAT 1080  
Db  
1223 GTTAAGACGTTGTTGATCAAGCGCTGACCAATTCGTTCTTGTAGTAACAGAAAGAT 1282  
Qy  
1081 GTGACTTTGATATTGATGACCAAAATTTGAAAACAGCGCTTATCATCCAGTTTATGGG 1140  
Db  
1283 GTGACTTTGATATTGATGACCAAAATTTGAAAACAGCGCTTATCATCCAGTTTATGGG 1342  
Qy  
1141 ATCTTGATCAATGGGCAACCAATTTGAAAATATCTGGAAGAAAGAACTGTTTACG 1200  
Db  
1343 ATCTTGATCAATGGGCAACCAATTTGAAAATATCTGGAAGAAAGAACTGTTTACG 1402  
Qy  
1201 TTTGATGTCCTAAAGTTGCGAATTTCTGTGACCAAGAACAGAAACCGGGAGACAGCA 1260  
Db  
1403 TTTGATGTCCTAAAGTTGCGAATTTCTGTGACCAAGAACAGAAACCGGGAGACAGCA 1462  
Qy  
1261 TGTGAGATTCCTGGATTT-----TGCTTAATGCTCCAGTGTAGTGTCAACT 1311  
Db  
1463 TGTGAGATTCCTGGATTTAAATGGAGAGTGCCTTAATGTTCCAGTGTAGTGTCAACT 1522  
Qy  
1312 CCAGCTCCCTGTATTTGCTCCTCGGAAACCCAGGACTTCAAGGCCCCCAAGGTGACCT 1371  
Db  
1523 CCAGCTCCCTGTATTTGCTCCTCGGAAACCCAGGACTTCAAGGCCCCCAAGGTGACCT 1582  
Qy  
1372 GSNCTGCTGGAAACCTGGCTACCTGGAACAACTGTTCAAGATGTTAAGCTGGATAT 1431  
Db  
1583 GSNCTGCTGGAAACCTGGCTACCTGGAACAACTGTTCAAGATGTTAAGCTGGATAT 1642  
Qy  
1432 CAGGGAATTCAGGAGACACAGGTGTTCCAGGATCTCCAGGAATAACAAGGAGCTCGAGA 1491  
Db  
1643 CAGGGAATTCAGGAGACACAGGTGTTCCAGGATCTCCAGGAATAACAAGGAGCTCGAGA 1702  
Qy  
1492 CTACAGGTTAACAAGGAACACAGGCGAGATGGTGAACAGGTTGATCGTGGACTTCT 1551  
Db  
1703 CTACAGGTTAACAAGGAACACAGGCGAGATGGTGAACAGGTTGATCGTGGACTTCT 1762  
Qy  
1552 GGTTTTCTGGGCTTCATGCGATCCAGGATCAAGGGTGAATGGTGGCCAAAGAGAC 1611  
Db  
1763 GGTTTTCTGGGCTTCATGCGATCCAGGATCAAGGGTGAATGGTGGCCAAAGAGAC 1822  
Qy  
1612 AAAGGATCACTGGATTTTATGCAAAAAGGTTGCAAAAGGTGAAAAGGGGAATGCTGGC 1671  
Db  
1823 AAAGGATCACTGGATTTTATGCAAAAAGGTTGCAAAAGGTGAAAAGGGGAATGCTGGC 1882  
Qy  
1672 TTCCTGGCTCCTGGACTCTGGAGAACAGGAGACATGGAAGAGATGGAATTAATG 1731  
Db  
1883 TTCCTGGCTCCTGGACTCTGGAGAACAGGAGACATGGAAGAGATGGAATTAATG 1942  
Qy  
1732 GGTAGTCCCGTTTCAAGGAGAGAGAGATCCCTGTTGCTCCCGGGCAGAGTGAACA 1791  
Db  
1943 GGTAGTCCCGTTTCAAGGAGAGAGAGATCCCTGTTGCTCCCGGGCAGAGTGAACA 2002  
Qy  
1792 CGGGAGAGCTCGAATTCAGGATTTCTGGAACCCAGGATTAATGGGCCAAAGGGA 1851  
|||||

2003 CGGCGAGAGCCTGGATCCAGGATTTCTCGAAAACAGAGATTATATGGGCCAAAAGGGA 2062  
Qy  
1852 GAAATTTGGCTCCTCAGGACAGCAAGGAAAAAAGAGACCCCAAGGATGCTGTTTAAATG 1911  
Db  
2063 GAAATTTGGCTCCTCAGGACAGCAAGGAAAAAAGAGACCCCAAGGATGCTGTTTAAATG 2122  
Qy  
1912 GGAAGCAATGCTCAGCAGGCGCCGCTGGAACACCGGATCTAAGGGAACAAAGGTGAA 1971  
Db  
2123 GGAAGCAATGCTCAGCAGGCGCCGCTGGAACACCGGATCTAAGGGAACAAAGGTGAA 2182  
Qy  
1972 CTTGGAATTCAGGAGATCCTCGGGCTTCTGGCTCAAGGGAGAACCCAGAGCAACGGGT 2031  
Db  
2182 CTTGGAATTCAGGAGATCCTCGGGCTTCTGGCTCAAGGGAGAACCCAGAGCAACGGGT 2242  
Qy  
2031 TCCCAGAGAACCCAGGATACATGGGTTTACCGGGATTCAGGGAACAAAGGGGAGCAAA 2091  
Db  
2243 TCCCAGAGAACCCAGGATACATGGGTTTACCGGGATTCAGGGAACAAAGGGGAGCAAA 2302  
Qy  
2092 GGAATTCAGGAGTGAAGGATTTTCAAGGCTCAAAAGGGAGAAATGGAACACAGGGAAT 2151  
Db  
2303 GGAATTCAGGAGTGAAGGATTTTCAAGGCTCAAAAGGGAGAAATGGAACACAGGGAAT 2362  
Qy  
2152 CCAGGCAACAGGAGTGAAGGATTTTCAAGGCTCAAAAGGGAGAAATGGAACACAGGGA 2211  
Db  
2363 CCAGGCAACAGGAGTGAAGGATTTTCAAGGCTCAAAAGGGAGAAATGGAACACAGGGA 2422  
Qy  
2212 GAACTTGTGTCGAGGTCCTTGGATCAAAAGGAGAAATCTGGGTTGGATGCTTGAATG 2271  
Db  
2423 GAACTTGTGTCGAGGTCCTTGGATCAAAAGGAGAAATCTGGGTTGGATGCTTGAATG 2482  
Qy  
2272 GGGCCCGCAGTCTTCAAGGGCAACCTGGGATCCAGGTCCTCAGGCAACCCAGGTTG 2331  
Db  
2483 GGGCCCGCAGTCTTCAAGGGCAACCTGGGATCCAGGTCCTCAGGCAACCCAGGTTG 2542  
Qy  
2332 GATGGAGACCCGAGAGAGTCTTCAAGCAATTTATTCAGCAAGTTTCCACAGATGA 2391  
Db  
2543 GATGGAGACCCGAGAGAGTCTTCAAGCAATTTATTCAGCAAGTTTCCACAGATGA 2602  
Qy  
2392 ATAAGAGCCAGTACCTCAGTCTTCAAGTGGAGAAATTAAGAAATTTGATTCATTC 2451  
Db  
2603 ATAAGAGCCAGTACCTCAGTCTTCAAGTGGAGAAATTAAGAAATTTGATTCATTC 2662  
Qy  
2452 CTGTCCTCAATGCTCCTCGGATTTCTGGGCACTGCTGCTGATGAGCCAGAGGT 2511  
Db  
2663 CTGTCCTCAATGCTCCTCGGATTTCTGGGCACTGCTGCTGATGAGCCAGAGGT 2722  
Qy  
2512 CCAGAGGATTAACCTGCTTGGCAGAGAGATGCTTCTGCTGATTTAGTGGTGTCCCT 2571  
Db  
2723 CCAGAGGATTAACCTGCTTGGCAGAGAGATGCTTCTGCTGATTTAGTGGTGTCCCT 2782  
Qy  
2572 GGAAGTTCAGGTTCAAGGATTTAAAGGCTTACAGGAAGAAATGGGAAAAAGGGAGC 2631  
Db  
2783 GGAAGTTCAGGTTCAAGGATTTAAAGGCTTACAGGAAGAAATGGGAAAAAGGGAGC 2842  
Qy  
2632 CAAGGTTTGGATTCCTGGAGAAACAGGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2691  
Db  
2843 CAAGGTTTGGATTCCTGGAGAAACAGGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2902  
Qy  
2692 CTTGGAATAAGCAAGAGAGTCTTCCAGGAGACCCAGGTCCTGCTGCTGCTGCTGCTGCTGCT 2751  
Db  
2903 CTTGGAATAAGCAAGAGAGTCTTCCAGGAGACCCAGGTCCTGCTGCTGCTGCTGCTGCTGCT 2962  
Qy  
2752 CATGGAACCTTGGATTCAGGAGAACAGGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2811  
Db  
2963 CATGGAACCTTGGATTCAGGAGAACAGGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3022  
Qy  
2812 TGTGTTAGTGAATTTGAGAGAGATCTTCAAGGAGAGATCTGCTGCTGCTGCTGCTGCTGCT 2865  
Db  
3023 TGTGTTAGTGAATTTGAGAGAGATCTTCAAGGAGAGATCTGCTGCTGCTGCTGCTGCTGCT 3076  
|||||

```
QY 1852 GAAATTGGCTCCAGCAGCAAGGAAAAAAGAGCCCGAGGATCGCTGGTTTAATG 1911
Db 2063 GAAATTGGCTCCAGCAGCAAGGAAAAAAGAGCCCGAGGATCGCTGGTTTAATG 2122
QY 1912 GAAGCAATGGCTCACAGGCCAGCTTGAAACACCGGATCTAAGGGAAGCAAAAGTGAA 1971
Db 2123 GAAGCAATGGCTCACAGGCCAGCTTGAAACACCGGATCTAAGGGAAGCAAAAGTGAA 2182
QY 1972 CTGGGAATTCAGGGATCGCTGGGCTTCTGGCTCAAGGAGAGACAGGACCGGT 2031
Db 2183 CTGGGAATTCAGGGATCGCTGGGCTTCTGGCTCAAGGAGAGACAGGACCGGT 2242
QY 2032 TCCCGAGGAAACAGGATACATGGGTTTACCGGGATTCAGGAAAAAAGGGGACAAA 2091
Db 2243 TCCCGAGGAAACAGGATACATGGGTTTACCGGGATTCAGGAAAAAAGGGGACAAA 2302
QY 2092 GGAATCAAGGTGAAAAAGGTATTGAGGTCAAAAGGAGAAAAATGGAAGCAGGAAAT 2151
Db 2303 GGAATCAAGGTGAAAAAGGTATTGAGGTCAAAAGGAGAAAAATGGAAGCAGGAAAT 2362
QY 2152 CAGGGCAACAGGAAATCAAGGCCATCATGGTCAAAAGGAGAGAGAGGTGAAAGGGA 2211
Db 2363 CAGGGCAACAGGAAATCAAGGCCATCATGGTCAAAAGGAGAGAGAGGTGAAAGGGA 2422
QY 2212 GAACCTGGTTCGAGGTGCCATTTGGATCAAAAGGAGAAATCTGGGTCGATGGCTTGATG 2271
Db 2423 GAACCTGGTTCGAGGTGCCATTTGGATCAAAAGGAGAAATCTGGGTCGATGGCTTGATG 2482
QY 2272 GGGCCGCGAGGCTTAAGGGGCAACCTGGGATCCAGGTCCTCAGGACCCCGAGTTG 2331
Db 2483 GGGCCGCGAGGCTTAAGGGGCAACCTGGGATCCAGGTCCTCAGGACCCCGAGTTG 2542
QY 2332 GATGGAAGCCCGAAGAGAGATTTTCAGAACAATTTATTCGACAGATTTGACAGATGA 2391
Db 2543 GATGGAAGCCCGAAGAGAGATTTTCAGAACAATTTATTCGACAGATTTGACAGATGA 2602
QY 2392 ATAAGAGCCAGTACAGTCTTACTTCAGAGTGAAGAAATTTAGAAATTTGATCATTCG 2451
Db 2603 ATAAGAGCCAGTACAGTCTTACTTCAGAGTGAAGAAATTTAGAAATTTGATCATTCG 2662
QY 2452 CTGTCCCAATGGCTCCCGGGTATTCTCGGCCACCTGGTCCGATAGGCCAGAGGT 2511
Db 2663 CTGTCCCAATGGCTCCCGGGTATTCTCGGCCACCTGGTCCGATAGGCCAGAGGT 2722
QY 2512 CCAGAGGATTAAGTGGTTGGCAGAGAGATGGTGTCTCGGATTAAGTGGTGTCCCT 2571
Db 2723 CCAGAGGATTAAGTGGTTGGCAGAGAGATGGTGTCTCGGATTAAGTGGTGTCCCT 2782
QY 2572 GGACGTCAGGTGTCAGAGGATTAAGAGGCTTACAGGAGAGAAATGGGGAAGAGGAGC 2631
Db 2783 GGACGTCAGGTGTCAGAGGATTAAGAGGCTTACAGGAGAGAAATGGGGAAGAGGAGC 2842
QY 2632 CAGGGTTGGGTATCTCGAGAAAGGTCTCTCGGTCCCTCGGTCAGAGGAGGCT 2691
Db 2843 CAGGGTTGGGTATCTCGAGAAAGGTCTCTCGGTCCCTCGGTCAGAGGAGGCT 2902
QY 2692 CTTGGATTAAGCAAGAGGTCTCCAGAGACCCAGGTCTCCCTGGCAAGATGGAGAC 2751
Db 2903 CTTGGATTAAGCAAGAGGTCTCCAGAGACCCAGGTCTCCCTGGCAAGATGGAGAC 2962
QY 2752 CATGGAACCTGGAAATCAAGGGCAACAGGCCCCCCCCAGGCATCTGGAGCCCATCACTA 2811
Db 2963 CATGGAACCTGGAAATCAAGGGCAACAGGCCCCCCCCAGGCATCTGGAGCCCATCACTA 3022
QY 2812 TGTTTTAGTGTAAATTCGCAAGAGATTCGTTTCAGAAAGGACCAAACTATTAG 2865
Db 3023 TGTTTTAGTGTAAATTCGCAAGAGATTCGTTTCAGAAAGGACCAAACTATTAG 3076
```

RESULT 3

AX086127

LOCUS

DEFINITION

Sequence 79 from Patent WO0112659.

linear

DNA

4160 bp

PAT 09-MAR-2001

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches 2861;

Conservative

C;

Mismatches

4;

Indels

9;

Gaps

1;

AX086127

AX086127.1

GI:13275978

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

Wiemann, S.

Human dna sequences

Patent: WO 0112659-A 79 22-FEB-2001;

German Human Genome Project (DE)

Location/Qualifiers

1..4160

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

1271 a

795 c

1026 g

1068 t

99.1%; Score 2839.6;

DB 6;

Length 4160;

99.5%; Pred. No. 0;

0;

<p>                     AVESTILYLGNTKTKGATQFALDYLFAKSSRLTKIAVILTDGSKSDQDVDAQAARD                      SKITLFAVGSEDEDAELRAIANKPSSTYVYVEDYIAISKREVMKQKLCBESVCP                      TRIPVAARDERGLDILGLDVKNKVKRIQLSPKIKGYEITSKVDLSLTSNVPFEG                      LPSPVVFVSTQRFKVKIWLRIILITIDGRPQIAVTLGVKILFTTTSVINGSOV                      TFANPQVTLFDFEGWHQIRLLVTEODVLYIDDOQIENKPLHVLGLINGQTOIGY                      SKEESTVDFDVKLRILYCDPENNRETAELPGNFECLNGSPSDVSTPAPCLPPCK                      PGLQPKPDPLPGLPNPGYPGQDQKGYGLAGTQVPSPIQANGLPGYKSEF                      GRDGRHGLPGLPGLHGMPSGEMKAGDKSPGYFKGANGKGNAGPFGJF                      PACEPRHGLDGLMSGPFKGEASPGAPGQDGTGEPIPGFPGNGLMGQKSEIGP                      PQQCKGKAGMPGLMGSNGSPGPTGSKSGKEBFIQMPGASGLKGEPCATGSP                      GFGVYGLPIQCKKGDIOGKEGKIOGQKENGROGIPGQOIGQHGKAKGERGK                      BFGVGAITSKESGVDLGMPAGKPGOPGDPGPGPLGKXKPRFSQFROVCT                      DVIRALPVLVLOSRIENCDHCLSOHSGSPIGPPGPIGPEGPRGLPLGRDGVPL                      VGVPRPGVRLGLPLGRNKGSGCFYRGEQGPPIGPEGPPGISKEGPRGDPG-                      PKGDGHPVRLGQGGPPIGCDPLCSFVIAKRDPRFKGNY"                 </p>									
BASE COUNT	1270	a	796	c	1026	g	1068	t	
ORIGIN									
Query March	99.2%	Score	2841.2	DB	9	Length	4160		
Best Local Similarity	99.6%	Pred. No. 0							
Matches 2862	Conservative	0	Mismatches	3	Indels	9	Gaps	1	
QY	1	ATGGCTCACTATATACATTTCTTGCATGCTTTTGGTCTCTTTCAGAACTTCGTG	60						
DB	203	ATGGCTCACTATATACATTTCTTGCATGCTTTTGGTCTCTTTCAGAACTTCGTG	262						
QY	61	TTAGCTGACAGTGGGAGCTACATCAGTTCCTCTACTGCTCCGACAGATTTAGTTTC	120						
DB	263	TTAGCTGACAGTGGGAGCTACATCAGTTCCTCTACTGCTCCGACAGATTTAGTTTC	322						
QY	121	ATCTTAGATGGCTCTTATAGTGTGGCCAGAAAACCTTTGAAATAGTGAAGAGTGGCTT	180						
DB	323	ATCTTAGATGGCTCTTATAGTGTGGCCAGAAAACCTTTGAAATAGTGAAGAGTGGCTT	382						
QY	181	GTCAATATCACAAAACCTTCACATAGGCGCGAAGTTTATTCAGTTGGAGTGGTCAA	240						
DB	383	GTCAATATCACAAAACCTTCACATAGGCGCGAAGTTTATTCAGTTGGAGTGGTCAA	442						
QY	241	TATAGTGACTACCTCTGCTGAGATTCCTCTCGAAGCTATGATTCAGGAGAACATTTG	300						
DB	443	TATAGTGACTACCTCTGCTGAGATTCCTCTCGAAGCTATGATTCAGGAGAACATTTG	502						
QY	301	ACGGCAGCAGTGGATTCCTACTCTTAGGAGGAAACAACAGACAGGAGGCGCATC	360						
DB	503	ACGGCAGCAGTGGATTCCTACTCTTAGGAGGAAACAACAGACAGGAGGCGCATC	562						
QY	361	CAGTTTGGCTCCGATTTACCTTTTGGCAAGTCTCTCAGATTTCTGACTAAGATACAGTG	420						
DB	563	CAGTTTGGCTCCGATTTACCTTTTGGCAAGTCTCTCAGATTTCTGACTAAGATACAGTG	622						
QY	421	GTACTTAGATGGCAAAATCCCAAGATGACGTGAGATGAGCTTCAAGCAGCAGAGAT	480						
DB	623	GTACTTAGATGGCAAAATCCCAAGATGACGTGAGATGAGCTTCAAGCAGCAGAGAT	682						
QY	481	AGTAAGATACATTTTGTCTATTGGTGTGTTTACAGAAACAGAGATGCCAGCTTAGA	540						
DB	683	AGTAAGATACATTTTGTCTATTGGTGTGTTTACAGAAACAGAGATGCCAGCTTAGA	742						
QY	541	GCTATTGCCAACAGCTTCTGCTACTATTATGTTTATGTGGAAGACTTATATGCAATA	600						
DB	743	GCTATTGCCAACAGCTTCTGCTACTATTATGTTTATGTGGAAGACTTATATGCAATA	802						
QY	601	TCCAAATAGGAGATGATGAGCAGAAACTTTGTGAGAACTTGTCTGTCCCAACAGA	660						
DB	803	TCCAAATAGGAGATGATGAGCAGAAACTTTGTGAGAACTTGTCTGTCCCAACAGA	862						
QY	661	ATTCAGTGGCAGCTCGTGTATGAAAGGGATTTGATATTCTTTTAGGTTTAGATGTAAT	720						
DB	863	ATTCAGTGGCAGCTCGTGTATGAAAGGGATTTGATATTCTTTTAGGTTTAGATGTAAT	922						
QY	721	AAAAAGTTTAAAGAAAGATACAGCTTTCCAAAAGAGATTAAGAGGATTAAGTAA	780						
DB	923	AAAAAGTTTAAAGAAAGATACAGCTTTCCAAAAGAGATTAAGAGGATTAAGTAA	982						

QY	781	TCAAAAGTGTGATTTATCAGAACTCAGAGCAATGTTTCCAGAGAGGTCTTCTCCATCA	840
DB	983	TCAAAAGTGTGATTTATCAGAACTCAGAGCAATGTTTCCAGAGAGGTCTTCTCCATCA	1042
QY	841	TATGTATTTGTCTTACTCAAGATTTAAAGTCAAGAAATTTGGGATTTATGGAGATA	900
DB	1043	TATGTATTTGTCTTACTCAAGATTTAAAGTCAAGAAATTTGGGATTTATGGAGATA	1102
QY	901	TTAACTATTGATGGAAGGCCCAAAATAGCACTTACCTTAAATGTTGGGCAAAATCTTA	960
DB	1103	TTAACTATTGATGGAAGGCCCAAAATAGCACTTACCTTAAATGTTGGGCAAAATCTTA	1162
QY	961	TTATTATACACACAGCGTAATTAATGCGTCAAGTGGTACCTTGGTCAACCTCA	1020
DB	1163	TTATTATACACACAGCGTAATTAATGCGTCAAGTGGTACCTTGGTCAACCTCA	1222
QY	1021	GTTAAAGAGCTTTGTTTGTATCAAGGGTGGCACCATAATGCTCTCTAGTAAACAGAACAGAT	1080
DB	1223	GTTAAAGAGCTTTGTTTGTATCAAGGGTGGCACCATAATGCTCTCTAGTAAACAGAACAGAT	1282
QY	1081	GTGACTTTGTATTTATGATGACCAAAATTTGAAAGCAAGCCCTTACATCCAGTTTATGGG	1140
DB	1283	GTGACTTTGTATTTATGATGACCAAAATTTGAAAGCAAGCCCTTACATCCAGTTTATGGG	1342
QY	1141	ATCTTGATCAATGGGCAAAACCCAAATTTGGAATAATTTCTGGAAGAGAGAACTTTTCA	1200
DB	1343	ATCTTGATCAATGGGCAAAACCCAAATTTGGAATAATTTCTGGAAGAGAGAACTTTTCA	1402
QY	1201	TTTGATGTCCTAAAGATTTGGAATTTCTGTTGACCAAGCAAGAAACAACCGGGAGACAGCA	1260
DB	1403	TTTGATGTCCTAAAGATTTGGAATTTCTGTTGACCAAGCAAGAAACAACCGGGAGACAGCA	1462
QY	1261	TGTGAGATTTCTTGGATTTT-----TGCCCTTAATGTCCTGATGTAGGTTTCAACT	1311
DB	1463	TGTGAGATTTCTTGGATTTTAAAGGAGTTCCTTAAATGTCCTGATGTAGGTTTCAACT	1522
QY	1312	CCAGCTTCCCTGTATTTGTCTCCGGGAAACCAAGGACTTCAAGGCCCAAGGTGACCTT	1371
DB	1523	CCAGCTTCCCTGTATTTGTCTCCGGGAAACCAAGGACTTCAAGGCCCAAGGTGACCTT	1582
QY	1372	GGACTGCTTGGAAACCTTGGCTACCTGGACCAACTGTTGATGTTAGCTTGGATAT	1431
DB	1583	GGACTGCTTGGAAACCTTGGCTACCTGGACCAACTGTTGATGTTAGCTTGGATAT	1642
QY	1432	CAGGAAATTCAGGAGACACACAGGTGTTCCAGGATCTCCAGGATACAAAGGAGCTCGAGGA	1491
DB	1643	CAGGAAATTCAGGAGACACACAGGTGTTCCAGGATCTCCAGGATACAAAGGAGCTCGAGGA	1702
QY	1492	CTACAGGTTTCAAGAGGAGAACAGGGGAGATGGTCAAGGGTGTATCGTGGACTTCTT	1551
DB	1703	CTACAGGTTTCAAGAGGAGAACAGGGGAGATGGTCAAGGGTGTATCGTGGACTTCTT	1762
QY	1552	GGTTTCTTGGGCTTTCATGGGATGCCAGATCAAGAGGTGAAATGGGTGCAAGAGAGAC	1611
DB	1763	GGTTTCTTGGGCTTTCATGGGATGCCAGATCAAGAGGTGAAATGGGTGCAAGAGAGAC	1822
QY	1612	AAAGGATCACCTGGATTTTATGGGCAAAAGGGTGCAGAGGTGAAAGGGGATGCTGGC	1671
DB	1823	AAAGGATCACCTGGATTTTATGGGCAAAAGGGTGCAGAGGTGAAAGGGGATGCTGGC	1882
QY	1672	TTCCCTGGCTTGGGACTTGTGGACCAAGCAGAGGAGATGAAAGGATGGATTTAATG	1731
DB	1883	TTCCCTGGCTTGGGACTTGTGGACCAAGCAGAGGAGATGAAAGGATGGATTTAATG	1842
QY	1732	GGTAGTCCCGGCTTTCAGGAGGAGACAGGATCCCTGCTGCTCCGGGAGAGATGGAACA	1791
DB	1943	GGTAGTCCCGGCTTTCAGGAGGAGACAGGATCCCTGCTGCTCCGGGAGAGATGGAACA	2002
QY	1792	CGGGAGAGCTTGAAGTCCAGGATTTCTTGGAAACCGAGGATTTAATGGGCAAAAGGGA	1851
DB	2003	CGGGAGAGCTTGAAGTCCAGGATTTCTTGGAAACCGAGGATTTAATGGGCAAAAGGGA	2062

QY	1321	TGTAATTGTCTCCGGGAAACCCAGGACTTCAAGGCCCCCAAGGTGACCCCTGGACGCT	1380
DB	1483	TGTAATTGTCTCCGGGAAACCCAGGACTTCAAGGCCCCCAAGGTGACCCCTGGACGCT	1542
QY	1381	GGGAACCCCTGGCTACCTCGACCACTCGTCAAGATGTAAGCTCGATATCAAGGAATT	1440
DB	1543	GGGAACCCCTGGCTACCTCGACCACTCGTCAAGATGTAAGCTCGATATCAAGGAATT	1602
QY	1441	GGGGACACCGAGGTGTTCCAGGATCTCAAGGAATCAAGGAGCTCAAGGATCAAGGT	1500
DB	1603	GGGGACACCGAGGTGTTCCAGGATCTCAAGGAATCAAGGAGCTCAAGGATCAAGGT	1662
QY	1501	TACAAAGGAGAACCGAGGAGATGTCGACAGGCTGATCTGACCTTCTCGGTTTCTCT	1560
DB	1663	TACAAAGGAGAACCGAGGAGATGTCGACAGGCTGATCTGACCTTCTCGGTTTCTCT	1722
QY	1561	GGGCTTCATGGCATGCCAGGATCAAGGGTGAATGGGTGCCAAGGAGACAAAGGATCA	1620
DB	1723	GGGCTTCATGGCATGCCAGGATCAAGGGTGAATGGGTGCCAAGGAGACAAAGGATCA	1782
QY	1621	CCTGGATTTTATGGCAAGAGGTCGCAAGGTGAAGGGGCAATGCTGGCTTCCCTGGC	1680
DB	1783	CCTGGATTTTATGGCAAGAGGTCGCAAGGTGAAGGGGCAATGCTGGCTTCCCTGGC	1842
QY	1681	CTCCCTGGACCTGCTGGAGAACCCAGGAGACATGGAAAGGATGGATTAAATGGGTAGTCCC	1740
DB	1843	CTCCCTGGACCTGCTGGAGAACCCAGGAGACATGGAAAGGATGGATTAAATGGGTAGTCCC	1902
QY	1741	GGTTTCAAGGGAGACGAGATCCCTGCTGTCGCGGCGAGATGGACACCGGGAGAG	1800
DB	1903	GGTTTCAAGGGAGACGAGATCCCTGCTGTCGCGGCGAGATGGACACCGGGAGAG	1862
QY	1801	CCTGGAAATCCAGGATTTCTCGAAACCCAGGATTAATGGGCCCAAGGGAGAAATGGG	1860
DB	1963	CCTGGAAATCCAGGATTTCTCGAAACCCAGGATTAATGGGCCCAAGGGAGAAATGGG	2022
QY	1861	CCTCCAGGACAGCAAGGAAAAAGAGCCCGAGGATGCTCGTTTAATGGGAGCAAT	1920
DB	2023	CCTCCAGGACAGCAAGGAAAAAGAGCCCGAGGATGCTCGTTTAATGGGAGCAAT	2082
QY	1921	GGCTCACAGGACAGCTCGAACCCGGGATCTAAGGGAGCAAGAGGTGAACCTGGAAAT	1980
DB	2083	GGCTCACAGGACAGCTCGAACCCGGGATCTAAGGGAGCAAGAGGTGAACCTGGAAAT	2142
QY	1981	CAAGGATGCTGGGCTTCTGGGCTCAAGGGAGAACCCAGGAGCAACCGGTTCCCGAGG	2040
DB	2143	CAAGGATGCTGGGCTTCTGGGCTCAAGGGAGAACCCAGGAGCAACCGGTTCCCGAGG	2202
QY	2041	GAACCGAGATACATGGGTTTACCCGGGATTCAGGAAAAAGGGGACAAAGGAAATCAA	2100
DB	2203	GAACCGAGATACATGGGTTTACCCGGGATTCAGGAAAAAGGGGACAAAGGAAATCAA	2262
QY	2101	GGTGAAGAGGATTTAGGGTCAAGGGAGAAAAATGGAGACAGGGAAATTCAGGGGCA	2160
DB	2263	GGTGAAGAGGATTTAGGGTCAAGGGAGAAAAATGGAGACAGGGAAATTCAGGGGCA	2322
QY	2161	CAGGGAAATTCAGGGCATCATGTCGCAAGGAGAGAGGTGAAAAGGGAGAACCTGGT	2220
DB	2323	CAGGGAAATTCAGGGCATCATGTCGCAAGGAGAGAGGTGAAAAGGGAGAACCTGGT	2382
QY	2221	GTCCGAGGTGCCATTTGATCAAAAGGAGAAATCTGGGTTGATGGCTGATCGGGCCGCA	2280
DB	2383	GTCCGAGGTGCCATTTGATCAAAAGGAGAAATCTGGGTTGATGGCTGATCGGGCCGCA	2442
QY	2281	GGTCTTAAGGGCAACCTGGGATCCAGGTCCTCAGGGACCCCGAGGTTTGGATGGGAG	2340
DB	2443	GGTCTTAAGGGCAACCTGGGATCCAGGTCCTCAGGGACCCCGAGGTTTGGATGGGAG	2502
QY	2341	CCCGGAGAGAGTTTTCAGACAAATTTATTCACAAAGTTTGACAGATGTAATAGAGCC	2400
DB	2503	CCCGGAGAGAGTTTTCAGACAAATTTATTCACAAAGTTTGACAGATGTAATAGAGCC	2562
QY	2401	CAGCTACCACTTACTTCAGAGTGAAGAAATAGAAATTTGATGATCTGCTGCCAA	2460
DB	2563	CAGCTACCACTTACTTCAGAGTGAAGAAATAGAAATTTGATGATCTGCTGCCAA	2622
QY	2461	CATGGCTCCCGGGTATTCTCGGCCACCTGGTCCGATAGGCCAGAGGGTCCAGAGGA	2520
DB	2623	CATGGCTCCCGGGTATTCTCGGCCACCTGGTCCGATAGGCCAGAGGGTCCAGAGGA	2682
QY	2521	TTACTGGTTTGCAGAGAGATGGTGTCTCTGGATAGTGGTGTCTCTGGACGTCA	2580
DB	2683	TTACTGGTTTGCAGAGAGATGGTGTCTCTGGATAGTGGTGTCTCTGGACGTCA	2742
QY	2581	GGTGTCAAGGATTTAAAGGCTTACACAGGAGAAATGGGGAAAAAGGGAGCCAAAGGTTT	2640
DB	2743	GGTGTCAAGGATTTAAAGGCTTACACAGGAGAAATGGGGAAAAAGGGAGCCAAAGGTTT	2802
QY	2641	GGGTATCTCGAGAAACAGGCTCTCTGGTCCCGCAGGTCCTCAGAGGGCCCTCTCGAAATA	2700
DB	2803	GGGTATCTCGAGAAACAGGCTCTCTGGTCCCGCAGGTCCTCAGAGGGCCCTCTCGAAATA	2862
QY	2701	AGCAAGAGAGGTCCTCCAGGAGACCCAGGCTCTCCCTGGCAAGATGGACCATGGAAA	2760
DB	2863	AGCAAGAGAGGTCCTCCAGGAGACCCAGGCTCTCCCTGGCAAGATGGACCATGGAAA	2922
QY	2761	CCTGGAAATCCAGGCAACCCAGGCCCCCAGGATCTCGACCCCATCTATCTTTAGT	2820
DB	2923	CCTGGAAATCCAGGCAACCCAGGCCCCCAGGATCTCGACCCCATCTATCTTTAGT	2982
QY	2821	GTAAATTCAGAGAGATCCCTTCAGAAAGGACCAAACTATTAG	2865
DB	2983	GTAAATTCAGAGAGATCCCTTCAGAAAGGACCAAACTATTAG	3027
RESULT 2			
AF414088			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
Homo sapiens (human)			
Homo sapiens			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
Fitzgerald, J. and Bateman, J.F.			
1 (bases 1 to 4160)			
A new FACIT of the collagen family: COL21A1			
FEBS Lett. 505 (2), 275-280 (2001)			
MEDLINE			
PUBMED			
11566190			
2 (bases 1 to 4160)			
Fitzgerald, J. and Bateman, J.F.			
Direct Submission			
Submitted (23-AUG-2001) Cell and Matrix Biology, Murdoch Children's			
Research Institute, Flemington Rd, Parkville, Vic 3052, Australia			
Location/Qualifiers			
1. 4160			
/organism="Homo sapiens"			
/mol_type="mRNA"			
/db_xref="taxon:9606"			
/chromosome="6"			
/map="6p12-11"			
1. 4160			
/gene="COL21A1"			
203..3076			
/gene="COL21A1"			
/note="FACIT collagen; similar to protein encoded by			
Genbank Accession Number AL136624"			
/codon_start=1			
/product="collagen XXI"			
/protein_id="AA10227.1"			
/db_xref="GI:15593270"			
/translation="MARVITFLCMVLLVLLVLLVLEDEGEVSSCRATPTDLVFLDGL			
SYSGPENFIVKWLNVITNFDFGPKFIQGVQVQVYPLVLEPLGSGYSGEHLTA			



Fri Oct 3 14:49:25 2003

```
collagen (COL21A1) gene
Genomics 79 (3), 395-401 (2002)
MEDLINE 21853310
PUBMED 11863369
REFERENCE 2 (bases 1 to 4122)
AUTHORS Chou,M.-Y.
TITLE Cloning and identification of a novel human alpha 1 chain-like
collagen (COLA1L) gene, a new member of the FACIT family
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 4122)
AUTHORS Chou,M.-Y.
TITLE Direct Submission
JOURNAL Submitted (20-DEC-2000) Biomedical Engineering Center, Industrial
Technology Research Institute, 195-85, Sec. 4, Chung Hsing Rd.,
Chutung, Hsinchu 310, Taiwan
FEATURES
    Location/Qualifiers
        1..4122
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /chromosome="6"
            /map="6p12.3-p11.2"
            /gene="COLA1L"
            163..3027
                /gene="COLA1L"
                /codon_start=1
                /product="alpha 1 chain-like collagen COLA1L precursor"
                /protein_id="AAU50033.1"
                /db_xref="GI:11974510"
                /translations="MAHYITFLCMLVLLQNSVLABGEVRSRRTAPDLVFIIDG
                SYSGENPEIVKRLNIITQNFIDIGPKFIQGVQVQSDYDPIVLEPLGSDSGEHLTA
                AVEISILVGGNTKGAIOFALDYLPAKSRFLTKIAVULIDGKSDQDDVDAQAARD
                SKITLPAIGVRSDEDAELRAIANKPSYVVFVDYIAISKIREVMKOLCESVCP
                TRIPAAORBERGDIILLGVNKKVKRIQLSPKKIKGVENTSKVDLSELTSNVPFEG
                LPSYVFTSTORFVKIKINDLRLLTIDRPOIAVTLNGLVKILLFTTSVINGSQVY
                IFANQVKTUFDGEGHQIRLVTEQDQVILYIDQIQENKPLHPVLGILINQIQIKY
                SKKEETVQDFQRIYCDPEQNNRETAICEIPFCLNGPSDVGTAPPCICPPKPGKL
                QPKPGPGLPKNPGYPOQPGDKPGYGIAGTGPVPGSPGIQAGARGLPYKGEPRD
                GDKGDRGLPFGPLHGMVPGSKGEMAKDGKPGFYGKKGAKGKNAGFPLGPGAG
                EPGRHGDKGLMGSPGFKBAGSPGAPQDQTRGEPGIPGFNRLGMLQKEIGIPPGQ
                QKKGAKRPMGLMWSNGSPQPGTSGSKGKEPGIQLQMPGASGLKGPBGATGSPGP
                GYMLPGLQKKGDKNGQSEKGIQGGKNGRGGIIPGGGIQGHKAGKBERGKEPG
                VRGAISKSGSDVGLMGPAGPKQPGDPGPGPGLDGLKPGRESSEGFIRQVCTEVI
                RAQLPVLLQSGRIKNDCHLSQHSFGIPGPPIPGPBPGLPGLPGRDGVPGPLGV
                FGRPGVRLGLPGRNKGSGQGFYGPGEQPPGPPGPGPGI1SKGFPFGDPLGPG
                DGDHKGPGIQQQPGPGICDPSLCFSVIAARRDPFRKPNY"
            163..228
                /gene="COLA1L"
                4072..4077
                    /gene="COLA1L"
BASE COUNT 1280 a 769 g 1077 t
ORIGIN
Query Match 100.0%; Score 2865; DB 9; Length 4122;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2865; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ATGCTCATATTTACATTTCTGCTGATGGTTTGGTCTGCTCTTTCAGAAATCTGTC 60
Db 163 ATGCTCATATTTACATTTCTGCTGATGGTTTGGTCTGCTCTTTCAGAAATCTGTC 222
Qy 61 TTAGCTGAAGTGGGGAAGTAAAGATCAAGTTGTGCTACTGCTCCGACAGATTAGTTTC 120
Db 223 TTAGCTGAAGTGGGGAAGTAAAGATCAAGTTGTGCTACTGCTCCGACAGATTAGTTTC 282
Qy 121 ATCTTAGTGGCTCTTATAGTTGGGCCGAGAAAACCTTTGAATAAGTGAAGTGGCTT 180
Db 283 ATCTTAGTGGCTCTTATAGTTGGGCCGAGAAAACCTTTGAATAAGTGAAGTGGCTT 342
Qy 181 GTCAATATCAAAAACCTTTGACATAGGCGCGAAGTTTATCAAGTTGAGTGGTTCAA 240
Db 343 GTCAATATCAAAAACCTTTGACATAGGCGCGAAGTTTATCAAGTTGAGTGGTTCAA 402
sig_peptide
polyA_signal
4072..4077
/gene="COLA1L"
BASE COUNT 1280 a 769 g 1077 t
ORIGIN
Qy 241 TATAGTGAATACCTGTGCTGGAGATTCCTCTCGAAGCTATATGATTCAGAGAACATTTG 300
Db 403 TATAGTGAATACCTGTGCTGGAGATTCCTCTCGAAGCTATATGATTCAGAGAACATTTG 462
Qy 301 AGGCGAGAGTGGGAATCCATCTACTTAGGAGGAAACACAAAGACAGAGGAGGCGCATC 360
Db 463 AGGCGAGAGTGGGAATCCATCTACTTAGGAGGAAACACAAAGACAGAGGAGGCGCATC 522
Qy 361 CAGTTTGGGCTCGATTACCTTTTTCGCAAGTCTCTCAGGATTTCTGACTAAGATAGAGTG 420
Db 523 CAGTTTGGGCTCGATTACCTTTTTCGCAAGTCTCTCAGGATTTCTGACTAAGATAGAGTG 582
Qy 421 GTACTTAGGATGGCAATCCCAAGATGAGTCAAGATGCAAGTCAAGTCAAGTCAAGAT 480
Db 583 GTACTTAGGATGGCAATCCCAAGATGAGTCAAGATGCAAGTCAAGTCAAGTCAAGAT 642
Qy 481 AGTAAAGATAACATTAATTGGCTATTGGTGTGTTTCAGAAAACAGAGATGCGCAACTAGA 540
Db 643 AGTAAAGATAACATTAATTGGCTATTGGTGTGTTTCAGAAAACAGAGATGCGCAACTAGA 702
Qy 541 GCTATTGCCAACAGCCTTCTGCTACTATGCTGTTTTATGTGAGAGACTATATTCGATA 600
Db 703 GCTATTGCCAACAGCCTTCTGCTACTATGCTGTTTTATGTGAGAGACTATATTCGATA 762
Qy 601 TCCAAAATAAGGGAAGTGAAGCAGAAAACCTTTGTGAAGAATCTGCTGTCGCAACACGA 660
Db 763 TCCAAAATAAGGGAAGTGAAGCAGAAAACCTTTGTGAAGAATCTGCTGTCGCAACACGA 822
Qy 661 ATTCAGTGGGAGCTGTGATGAAGGGGATTTGATATCTTTAGCTTTAGTGTAAAT 720
Db 823 ATTCAGTGGGAGCTGTGATGAAGGGGATTTGATATCTTTTAGGTTTAGATGTAAAT 882
Qy 721 AAAAAAGTTAAAGAAAGATACAGCTTTCCACAAAAGATATAAGAGATATGAATGAACA 780
Db 883 AAAAAAGTTAAAGAAAGATACAGCTTTCCACAAAAGATATAAGAGATATGAATGAACA 942
Qy 781 TCAAAAGTTGATTTATCAGAACTCAAGAAAGATCTTTTCCAGAGGCTCTTCTCCATCA 840
Db 943 TCAAAAGTTGATTTATCAGAACTCAAGAAAGATCTTTTCCAGAGGCTCTTCTCCATCA 1002
Qy 841 TATGATTTGTCTACTCAAGAGTTTAAAGTCAAGAAAATTTGGGATTTATGGGAAATA 900
Db 1003 TATGATTTGTCTACTCAAGAGTTTAAAGTCAAGAAAATTTGGGATTTATGGGAAATA 1062
Qy 901 TTAACCTATTGATGGAGGCGCAAAATAGCAGTTACTTTAAATGCTGTGGACAAATCTTA 960
Db 1063 TTAACCTATTGATGGAGGCGCAAAATAGCAGTTACTTTAAATGCTGTGGACAAATCTTA 1122
Qy 961 TTATTTACAAACACAGCGTAAATTAATGCTCACAAGTGGTTACCTTTGTAACCTCAA 1020
Db 1123 TTATTTACAAACACAGCGTAAATTAATGCTCACAAGTGGTTACCTTTGTAACCTCAA 1182
Qy 1021 GTTAAGACCTGTGTTGATGAAGGCTGGCAACAAATTCGCTCTCTAGTAAACAGAGAT 1080
Db 1183 GTTAAGACCTGTGTTGATGAAGGCTGGCAACAAATTCGCTCTCTAGTAAACAGAGAT 1242
Qy 1081 GTGACTTTGATATTGATGAGCCCAAAATTCGCTCTCTAGTAAACAGAGATTTAGGG 1140
Db 1243 GTGACTTTGATATTGATGAGCCCAAAATTCGCTCTCTAGTAAACAGAGATTTAGGG 1302
Qy 1141 ATCTTAGTCAATGGGCAACCCAAATTCGAAAAATATTCTGGAAAAAGAAAGTGTTCAG 1200
Db 1303 ATCTTAGTCAATGGGCAACCCAAATTCGAAAAATATTCTGGAAAAAGAAAGTGTTCAG 1362
Qy 1201 TTTCATGTCCTCAAAAGTTGCGAATCTACTGTGACCCAGACAGACACCGGAGACAGCA 1260
Db 1363 TTTCATGTCCTCAAAAGTTGCGAATCTACTGTGACCCAGACAGACACCGGAGACAGCA 1422
Qy 1261 TGTGAGATTCCTGGATTTTGGCTTAATGCTCCAGTGTATGTAGTTCGCTCAAGTCTCC 1320
Db 1423 TGTGAGATTCCTGGATTTTGGCTTAATGCTCCAGTGTATGTAGTTCGCTCAAGTCTCC 1482
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2003, 18:36:28 ; Search time 10133 Seconds  
(without alignments)  
11566.770 Million cell updates/sec

Title: US-09-996-611b-5  
Perfect score: 2865  
Sequence: 1 atggctcactattacatt.....gaaaggaaccaactattag 2865

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl.\*  
1: gb.ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.ro.\*  
11: gb.sts.\*  
12: gb.sy.\*  
13: gb.un.\*  
14: gb.vi.\*  
15: em.ba.\*  
16: em.fun.\*  
17: em.hum.\*  
18: em.in.\*  
19: em.mu.\*  
20: em.om.\*  
21: em.or.\*  
22: em.ov.\*  
23: em.pat.\*  
24: em.ph.\*  
25: em.pl.\*  
26: em.ro.\*  
27: em.sts.\*  
28: em.un.\*  
29: em.vi.\*  
30: em.htg.hum.\*  
31: em.htg.inv.\*  
32: em.htg.other.\*  
33: em.htg.mus.\*  
34: em.htg.pln.\*  
35: em.htg.rod.\*  
36: em.htg.mam.\*  
37: em.htg.vrt.\*  
38: em.sy.\*  
39: em.htgo.hum.\*  
40: em.htgo.mus.\*  
41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	2865	100.0	4122	9	AF330693	AF330693 Homo sapi
2	2841.2	99.2	4120	9	AF414088	AF414088 Homo sapi
3	2839.6	99.1	4160	9	AX086127	AX086127 Sequence
4	2839.6	99.1	4160	9	HSM801537	ALL136624 Homo sapi
5	1648	36.6	2295	9	AK096444	AK096444 Homo sapi
6	974	34.0	1184	6	AX430410	AX430410 Sequence
7	972.4	33.9	1169	6	AX430281	AX430281 Sequence
8	908	31.7	1125	6	AX552238	AX552238 Sequence
9	553.6	19.3	110000	9	AF438327_2	Continuation (3 of
10	553.6	19.3	126501	9	HS708F5	AL031782 Human DNA
11	453.6	15.8	1414	6	AX721209	AX721209 Sequence
12	353.4	12.3	447	6	BD058397	BD058397 Secrete
13	303	10.6	3164	5	CHKGSCND	K01702 Chicken cho
14	295.4	10.3	2927	9	HSCOL9AL	X54412 Human mRN
15	292.6	10.2	2359	3	AK115182	AK115182 Clona int
16	289.6	10.1	2766	10	MJSCOLA1IX	L12215 Mouse colla
17	289.6	10.1	3049	10	MUSNA411	D17511 Mus musculu
18	285	9.9	3619	9	BC042075	BC042075 Homo sapi
19	285	9.9	6352	9	AF406780	AF406780 Homo sapi
20	282.6	9.9	3178	5	AY113700	AY113700 Dario rer
21	279	9.7	2388	10	AF349718	AF349718 Mus muscu
22	276.4	9.6	3074	5	BC045244	BC045244 Xenopus l
23	272.4	9.5	2416	5	CHK3ACOL	M83179 Gallus gall
24	272.2	9.5	2901	5	GOCOLA3IX	X64712 G.gallus mR
25	268.2	9.4	4428	6	AX146430	AX146430 Sequence
26	263.2	9.2	3720	9	SF57132	SF57132 COL16A1=typ
27	262.8	9.2	4473	3	AF525468	AF525468 Hydra vul
28	262	9.1	3044	5	BC046861	BC046861 Xenopus l
29	261.6	9.1	5387	9	HJMCOL16A	X92642 Hmo sapien
30	261.6	9.1	5387	11	G28574	G28574 human STS S
31	259.8	9.1	4428	6	AX146422	AX146422 Sequence
32	259.8	9.1	4428	6	AX146424	AX146424 Sequence
33	259.6	9.1	3902	9	BC028178	BC028178 Homo sapi
34	251	8.8	3967	10	BC028248	BC028248 Mus muscu
35	251	8.8	4778	10	BC043089	BC043089 Mus muscu
36	251	8.8	4786	10	BC052398	BC052398 Mus muscu
37	250.2	8.7	2777	5	SGU07973	U07973 Gallus gall
38	248.8	8.7	1955	10	BC023940	BC023940 Mus muscu
39	248.8	8.7	5160	10	BC027766	BC027766 Mus muscu
40	248.2	8.7	2263	4	CFU07889	U07888 Canis fami
41	248	8.7	5646	10	AB041350	AB041350 Mus muscu
42	247	8.6	5909	3	NEMA2C4A	M67507 A.suum alph
43	246.4	8.6	4235	10	BC016479	BC016479 Mus muscu
44	245.4	8.6	2667	5	CHKCOLA1XI	M88593 Gallus gall
45	242.6	8.5	5460	6	AX302553	AX302553 Sequence

# ALIGNMENTS

RESULT :  
AF330693  
LOCUS  
DEFINITION Homo sapiens alpha 1 chain-like collagen COL1A1 precursor (COL1A1)  
mRNA, complete cds.  
ACCESSION AF330693  
VERSION AF330693.1 GI:17974509  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 4122)  
AUTHORS Chou,M.Y. and Li,H.C.  
TITLE Genomic organization and characterization of the human type XXI